

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:24 ; Search time 58.6787 Seconds  
(without alignments)  
2176.455 Million cell updates/sec

Title: US-09-857-518A-6  
Perfect score: 2173  
Sequence: 1 MEXIEVSIKSTSS.....MAMLEQDPFLALAPKILI 452

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_29Jan04.\*
- 1: Geneseq1980s.\*
  - 2: Geneseq1990s.\*
  - 3: Geneseq2000s.\*
  - 4: Geneseq2001s.\*
  - 5: Geneseq2002s.\*
  - 6: Geneseq2003as.\*
  - 7: Geneseq2003bs.\*
  - 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2373	100.0	452	3	AAB36439	Strawberr
2	2373	100.0	452	3	AAY79656	Strawberr
3	2043	86.1	455	3	AAB36454	Strawberr
4	538.5	22.7	726	3	AAG39696	Arabidops
5	538.5	22.7	823	3	AAG39695	Arabidops
6	538.5	22.7	954	3	AAG39694	Arabidops
7	533.5	22.5	726	3	AAG27611	Arabidops
8	533.5	22.5	823	3	AAG27610	Arabidops
9	533.5	22.5	954	3	AAG27609	Arabidops
10	499	21.0	426	3	AAB36458	Lemon alc
11	498.5	21.0	443	5	ABB92488	Herbicida
12	498	21.0	436	5	ABB91119	Herbicida
13	498	21.0	436	7	ADB91973	Acyltrans
14	495	20.4	446	7	ADB91985	Acyltrans
15	448	18.9	435	7	ADB91986	Acyltrans
16	438	18.5	474	6	ABB82998	P. somnif
17	438	18.5	474	6	ABB82991	P. somnif
18	438	18.5	474	6	ABB82997	P. somnif
19	408.5	17.2	439	2	AAY17403	Clarkia c
20	404.5	17.0	433	2	AAY17400	Clarkia b
21	404.5	17.0	433	6	AAE29783	C. brewer
22	404.5	17.0	433	6	ABB82993	C. brewer
23	404.5	17.0	433	7	ADB91977	Acyltrans
24	397.5	16.8	439	2	AAY17402	Clarkia c
25	387.5	16.3	439	3	AAY51247	C. roseus

26	387.5	16.3	439	6	ABB82992	C. roseus
27	387.5	16.3	439	7	ADB91979	Acyltrans
28	386	16.3	457	4	AAY08789	Petunia a
29	386	16.3	460	4	AAY08790	Nierember
30	385	16.2	408	4	AAY08791	Black Pan
31	384	16.2	438	2	AAY17401	Clarkia c
32	377	15.9	406	4	AAY08792	Black Pan
33	374	15.8	403	4	AAY08793	Light blu
34	357.5	15.1	393	4	AAY08794	Light blu
35	316.5	13.3	297	2	AAY17404	Clarkia c
36	308	13.0	331	5	ABB91120	Herbicida
37	308	13.0	331	7	ADB91972	Acyltrans
38	255	10.7	428	3	AAG24209	Arabidops
39	255	10.7	440	3	AAG24208	Arabidops
40	253	10.7	426	3	AAG37735	Arabidops
41	253	10.7	426	3	AAG37720	Arabidops
42	253	10.7	426	5	ABB93930	Herbicida
43	253	10.7	438	3	AAG37734	Arabidops
44	253	10.7	442	3	AAG37719	Arabidops
45	243.5	10.3	436	6	ABU58179	Wheat str

ALIGNMENTS

RESULT 1  
AAB36439 standard; protein; 452 AA.  
XX AAB36439;  
XX DT 28-FEB-2001 (first entry)  
XX DE Strawberry alcohol acyl transferase protein SEQ ID NO:13.  
XX KW Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase;  
KW aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;  
KW pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;  
KW alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;  
KW food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;  
KW confectionery; flavouring; oral medication; vitamin; aroma; beverage;  
KW alcohol; scent; fragrance; perfume; cosmetic; suspension aid;  
KW aluminium salt; anti-perspirant; pharmaceutical; cleaning product;  
KW insect pheromone; dye carrier; solvent; insect repellent; miticide;  
KW scabicide; plasticiser; deodorant.  
XX OS Fragaria x ananassa.  
XX PN WO200032789-A1.  
XX PD 08-JUN-2000.  
XX PF 02-DEC-1999; 99WO-NL000737.  
XX PR 02-DEC-1998; 98EP-00204018.  
XX PR 12-MAR-1999; 99EP-00200739.  
XX (CPRO-) CPRO-DLO CENT PLANTENVERDELINGS REPROD.  
XX Aaroni A, Luecker J, Verhoeven HA, Van Tunen AJ, O'connell AP;  
N-PSDB; AAC64766.  
XX WPI; 2000-412335/35.  
XX A new DNA sequence encoding a polypeptide with alcohol acyl transferase  
activity for producing and regulating aromatic and/or aliphatic ester  
formation in microorganisms, plant cells or plants.  
XX Claim 7; Page 71; 163pp; English.  
XX The present invention describes nucleotide sequences with thiolase,  
alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
aminotransferase and esterase activities, which are involved in the



QY 301 NLFWAQALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYLD 360  
 DB 301 NLFWAQALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYLD 360  
 QY 361 FORTWSSMEPADDIYLFSSWTFNFPDQGWRTSWIGVAGKIESASCKFIILVPTQCGS 420  
 DB 361 FORTWSSMEPADDIYLFSSWTFNFPDQGWRTSWIGVAGKIESASCKFIILVPTQCGS 420  
 QY 421 GTEAVNLEEKWAMLEQDPHELALASPKTLI 452  
 DB 421 GTEAVNLEEKWAMLEQDPHELALASPKTLI 452

## RESULT 3

AAAB36454  
 ID AAB36454 standard; protein; 455 AA.

AC AAB36454;

DT 28-FEB-2001 (first entry)

Strawberry vesca alcohol acyl transferase protein SEQ ID NO:168.

Strawberry: fruit flavour; biosynthetic pathway; aliphatic; thiolase;  
 aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;  
 pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;  
 alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;  
 food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;  
 confectionery; flavouring; oral medication; vitamin; aroma; beverage;  
 alcohol; scent; fragrance; perfume; cosmetic; suspension aid;  
 aluminium salt; anti-perspirant; pharmaceutical; cleaning product;  
 insect pheromone; dye carrier; solvent; insect repellent; miticide;  
 scabicide; plasticiser; deodorant.  
 Fragaria x ananassa.

OS WO200032789-A1.

FN 08-JUN-2000.

PD 02-DEC-1999; 99WO-NL000737.

PF 02-DEC-1998; 98EP-00204018.

PR 12-MAR-1999; 99EP-00200739.

XX (CPRO-) CPRO-DLO CENT PLANTENVEREDELINGS REPROD.

XX Aharoni A, Luecker J, Verhoeven HA, Van Tunen AJ, O'connell AP;

XX WPI; 2000-412335/35.

DR N-PSDB; AAC64750.

XX A new DNA sequence encoding a polypeptide with alcohol acyl transferase  
 PT activity for producing and regulating aromatic and/or aliphatic ester  
 PT formation in microorganisms, plant cells or plants.

XX Example 5; Page 107; 163pp; English.

XX The present invention describes nucleotide sequences with thiolase,  
 CC alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
 CC aminotransferase and esterase activities, which are involved in the  
 CC biosynthetic pathway for aliphatic and/or aromatic ester production in  
 CC fruit. The nucleotide sequences can be inserted into the genome of a  
 CC fruit-producing plant to regulate aliphatic and/or aromatic ester  
 CC formation. Aromatic and/or aliphatic esters in microorganisms, plant  
 CC cells or plants are produced by inserting thiolase, alcohol acyl  
 CC transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
 CC aminotransferase and esterase nucleotide sequences into the genome and  
 CC feeding the microorganism or plant with alcohol, aldehydes, alpha-keto  
 CC acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and  
 CC their proteins can be used in the processed food industry as food  
 CC additives to enhance the flavour of syrups, ice-creams, frozen desserts,  
 CC yoghurts and confectionery. They are used: as flavouring agents for oral

CC medications and vitamins; provide flavour and aroma in beverages,  
 CC including alcohol; enhance or reduce fruit flavour, aroma, fragrance or  
 CC scent; enhance the flavour or aroma of natural, synthetic or artificial  
 CC products; for the production of novel combinations of artificial flavour  
 CC substances; as antibacterial or anti-fungal agents; as fragrance or  
 CC perfumes in cosmetics, creams, sun-protectant products, hair  
 CC conditioners, lengthening agents and fixatives in perfumes, cleaning  
 CC aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning  
 CC products, personal care products and animal care products; as  
 CC disinfectant additives; as degreasing solvents for electronics; as insect  
 CC pheromones; and as dye carriers, solvents, insect repellents, miticides,  
 CC scabicides, plasticisers and deodorants. The present sequence represents  
 CC strawberry vesca alcohol acyl transferase, from the present invention  
 XX

SQ Sequence 455 AA;

Query Match 86.1%; Score 2043; DB 3; Length 455;

Best Local Similarity 87.6%; Pred. No. 7.1e-203;

Matches 397; Conservative 15; Mismatches 39; Indels 2; Gaps 2;

QY 1 MEKIEVSINSKTIKPTSTSTPQPKYKLTLLDQLTTPAYVPIVFYFETDHD-FNLPTQL 59

DB 1 MEKIEVSIISKTIKPTSTSSPLQPKYKLTLLDQLTTPSYVVMVFYFETGPAVFNL-QTL 59

QY 60 ADLRQALSETLTLYPLSGRVKNLYIDDFEGVPYLEARVNCMDTDFLRKIECINEF 119

DB 60 ADLRHALSETLTLYPLSGRVKNLYIDDFEGVPYLEARVNCMDTDFLRKIECINEF 119

QY 120 VPKPFSMEAISDERYPPLLGQVNWETSGIAGVSVSHKLLIDGTADCFKSKGAVPRGC 179

DB 120 VPKPFSMEAISDERYPPLLGQVNIENSGIAGVSVSHKLLIDGTSDCFKSKCAVFRGS 179

QY 180 RENIHPISLEAALPPRPDIPEKYVDOMEALWFAKKVATRRFVGKAISSIODEAK 239

DB 180 RDKIHPNLSQALPPRPDDELPKVARQMEGLWFGKKVATRRFVGKAISSIODEAK 239

QY 240 SRSVPKPSRVHVTGFLWKHLIAASRALTSGLTSTRLSIAAQAVNLRNRMETVLDNAT 299

DB 240 SRSVPKPSRVQVTSFLNKLHATSRLTSGTSTRLSIATQVNISSRNMTVMDNAI 299

QY 300 GNLFWAQAILELSHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYL 359

DB 300 GNLFWAPALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYL 359

QY 360 DFQRTWSSMEPADDIYLFSSWTFNFPDQGWRTSWIGVAGKIESASCKFIILVPTQCG 419

DB 360 DFQRTWSSMEPAPEIYLF-TSWTNFFNQDQGWRTSWIGVAGKIESAFCLNLTILVPTPCD 419

QY 420 SGTEAVNLEEKWAMLEQDPHELALASPKTLI 452

DB 420 TG-EAVNLEEKWAMLEQDPQFLALASPKTLI 452

## RESULT 4

AAAG39696

ID AAG39696 standard; protein; 726 AA.

XX AAG39696;

AC AAG39696;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 49156.

DE Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.  
XX 99US-0121825P.  
PR 25-FEB-1999; 99US-0123180P.  
PR 03-MAR-1999; 99US-0123548P.  
PR 09-MAR-1999; 99US-0125788P.  
PR 23-MAR-1999; 99US-0126264P.  
PR 25-MAR-1999; 99US-0126785P.  
PR 29-MAR-1999; 99US-0127462P.  
PR 01-APR-1999; 99US-0128234P.  
PR 06-APR-1999; 99US-0128714P.  
PR 08-APR-1999; 99US-0129845P.  
PR 16-APR-1999; 99US-0130077P.  
PR 19-APR-1999; 99US-0130510P.  
PR 21-APR-1999; 99US-0130891P.  
PR 23-APR-1999; 99US-0131449P.  
PR 28-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139482P.  
PR 18-JUN-1999; 99US-0139484P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142801P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147301P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153756P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155113P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155559P.

```
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157111P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161982P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      22.7%  Score 538.5;  DB 3;  Length 726;
Best Local Similarity 33.6%;  Pred. No. 3.7e-46;
Matches 153;  Conservative 88;  Mismatches 178;  Indels 37;  Gaps 15;

QY 3 KIEVINSKHTIKPSTSTP--LQPYKLTLLDQTPPAYVPIVFYPTTDHDFNLPTOLA 60
DB 284 QMEVDVVSRIILKPS-SPTNHLKKFKLSLEQLGPIFGPWVFPYS-ANNSIKPTQLQ 341
QY 61 DLQALGETLTYPLSGRVNNLIYDDFEGVPILEARVNCMDTDFLRKIKICLNEFV 120
DB 342 MLKXSLSETLTHFYPLAGRLKGNISICNDSGADFLKARNVSPLSNLLLEPSSDSLQOLI 401
QY 121 PIKPFMEALSDERYPLGVQNVFDSG-TAIGVSVSHKLIDGTADCFKLSGAV-FRG 178
DB 402 PTVSDSIET-----RTRLLAQASFEGCSWISGICISHKLADATSIGLFMKSWAISR 457
QY 179 CRENIHPSLSEALPPP---RDDLPKYVDQMEALWFAGKKVATRRFVFGKAISSIQ 235
DB 458 SIKTIGAP-VFDIVKIPPGNPFSTSPAPVVEPIWM-----NQTLKRFIDSSSIQALQ 512
QY 236 DEAKSEVPKPSRVHATGFLWKHLIAASRALTSGTSTRLSIAAQVNLRTRNMETVL 295
DB 513 AKASSEVWQPTRVHVASALIWKSAMKATRTV-SGT--SKPSILANSVLSRVS-PPFT 568
QY 296 DNATKGLFWAQLLELSHITPETISDLKLDULNLLGSKQNGDYFEFTFKGEGYGR- 354
DB 569 KNSIGNLVSTPAKAE-----EGINQTKLOTLVSKIRKAKQRFDIHPKLVGNPNATEI 623
QY 355 MCBYLDQRTMSMEPAP--DIYLFSSWTFN-FNELDFMGRTSWIGVAGKIESACKFI 411
DB 624 ICXY---QKEAGDMIASGDDFDYIFSSACREGLYETDFGKPVWVGFPPVRQK---NIV 677
QY 412 ILVPTQCGSGIEAWNLLEBKAMLEQDPHPLAAS 447
DB 678 TLIDTTEAGGIEAWNLNEQEMNLFQDRELLQFAS 713
```

RESULT 5  
AAG39695  
ID AAG39695 standard; protein; 823 AA.  
XX  
AC AAG39695;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49155.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
DN EP1033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0137282P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.

```
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139898P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.

PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.7%; Score 538.5; DB 3; Length 823;
Best Local Similarity 33.6%; Pred. No. 4.5e-46;
Matches 153; Conservative 88; Mismatches 178; Indels 37; Gaps 18;

QY 3 KIEVSINSKHTIKESTSTP--LQPYKLCGLDQUTPPAYVPIVFPYITDHFNLPTQLA 60
Db 381 QMRVDVVSRDIKES-SPTPNELKKFKLSLLEQLGPTIFGEMVFYS-ANNSIKPTEQLQ 438
QY 61 DLROALSETLTYPLDSCRWNNLYIDDFEGVYLEARVNCMTDFLRLKIECLNFV 120
Db 439 MLKSLSETLTHFYPLAGLAKENISIDCNDSGADFLARVNSPLSNLLEPSSDLOQLI 498
QY 121 PIKPFMSAISCDERYPLLGQVQVVDGSG-IAIGVSVSHKLDGGTADCFKLSWGAV-FRG 178
Db 499 PTVSDSIET-----RTRLLAQSPFCGSMISGVCISHKLADATSIGLFMKSNAAISSRG 554
```



```
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150894P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155559P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0158232P.
PR 04-OCT-1999; 99US-0158369P.
PR 04-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150894P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155559P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0158232P.
PR 04-OCT-1999; 99US-0158369P.
PR 04-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

Query Match 22.7%; Score 538.5; DB 3; Length 954;
Best Local Similarity 33.6%; Pred. No. 5,7e-46;
Matches 153; Conservative 88; Mismatches 178; Indels 37; Gaps 18;

Qy 3 KIEVSINSKHTIKPSTSTP--LQPYKLTLLDLPYVPIVFEYPIVTOHDFNLQTLA 60
Db 512 QMRVDVVSRDIIKPS-SPTPNHLKKFKLSLLEQGTIFGPMVFYIS-ANKSIKPTQLQ 569
Qy 61 DLROALSETLTVYPLSGRVKNLYIDDFEGVYLEARNVCOMTDLELEKICLNEFV 120
Db 570 MLKKSLSLTHFYPLAGRLKGNISIDCNSGDGDFLEARNVSNLLEPSSLSQLQLI 629
Qy 121 PIKPFMEALSDERYPLLGQVNVDSG-IAIGVSVSHKLIDGTADCFELKSGAV-FRG 178
Db 630 PTVSDSIET---RTRLLLAQASFEGCSMSIGVICISHKLADA--SIGLEMKSWAISR 685
Qy 179 CRENIHPSLSSEALLPPP---RDLPKXYDQMEALWFAKKVATRVFVGVKAISIQ 235
Db 686 SIKTIGAP-VFDTVKIFPPGNFSETSPVVPPEIMK---NOTLSRRFIIDSSSIQALQ 740
Qy 236 DEAKSESVKPSRVHVTGFLWKHLIAASRALTSQTTSTRLSIAAQAQVNLKRMNMTVL 295
Db 741 AKASSFEVNPQTRVEAVSALIKSAMKATRV-SGT--SKESILANSVLSRSVY-PPFT 796
Qy 296 DNATGNLFWAQAILESHHTPEISDLKDLVNLKNGVKQNGDYFETPKGEGYGR- 354
Db 797 KNSIGNLVSYFAAKAE-----EGINOTKLQTLVSKIRKAKORFDDIHPKLVGNPNATEI 851
Qy 355 MCEYLDFORTMSSMEPAP--DIVLFSSWTF-FNPLDPGMRGRTSWIGVAGKIESASCKFI 411
Db 852 ICSTY---QKAGDMIASGDFDFYIFSSACRFLGYETDFGKGPVWVGFPSPVRQK---NIV 905
Qy 412 ILVPTOCGSGIEAWVNLSEKQAMLEQDPHPLALAS 447
Db 906 TLDDTKGAGGIEAWVNLSEKQAMLEQDPHPLALAS 941

RESULT 7
AAG27611
ID AAG27611 standard; protein; 726 AA.
XX
AC AAG27611;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 32516.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 25-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
```

PR 08-APR-1999; 99US-0128714P.  
 PR 16-APR-1999; 99US-0129845P.  
 PR 19-APR-1999; 99US-0130077P.  
 PR 21-APR-1999; 99US-0130449P.  
 PR 23-APR-1999; 99US-0130510P.  
 PR 28-APR-1999; 99US-0130891P.  
 PR 30-APR-1999; 99US-0131449P.  
 PR 30-APR-1999; 99US-0132048P.  
 PR 30-APR-1999; 99US-0132407P.  
 PR 04-MAY-1999; 99US-0132484P.  
 PR 05-MAY-1999; 99US-0132485P.  
 PR 06-MAY-1999; 99US-0132486P.  
 PR 06-MAY-1999; 99US-0132487P.  
 PR 07-MAY-1999; 99US-0132863P.  
 PR 11-MAY-1999; 99US-0134215P.  
 PR 14-MAY-1999; 99US-0134218P.  
 PR 14-MAY-1999; 99US-0134219P.  
 PR 14-MAY-1999; 99US-0134221P.  
 PR 14-MAY-1999; 99US-0134370P.  
 PR 16-MAY-1999; 99US-0134768P.  
 PR 19-MAY-1999; 99US-0134941P.  
 PR 20-MAY-1999; 99US-0135124P.  
 PR 21-MAY-1999; 99US-0135353P.  
 PR 24-MAY-1999; 99US-0135629P.  
 PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137528P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145132P.  
 PR 23-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 27-JUL-1999; 99US-0145919P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 09-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149932P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.

```
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159337P.
PR 14-OCT-1999; 99US-0159363P.
PR 14-OCT-1999; 99US-0159388P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.5%; Score 533.5; DB 3; Length 726;
Best Local Similarity 33.3%; Pred. No. 1.2e-45;
Matches 152; Conservative 87; Mismatches 180; Indels 37; Gaps 18;

Qy 3 KIEYSINKHTIKESTSTP--LOPYKLTLDLTTPPAYVIVPYPITDHDNLPQVLA 60
Db 284 QMRVDVVSRODIKES-SPTNHLKKFKLSLEQLGPTIFGMPFFYS-ANNSIKPTQLQ 341
Qy 61 DLQALSETLLYPLSGRVKNLYIDFEGVPYLKRVNCDMTDFLRKIKIECLNEFV 120
Db 342 MLKKSLETLTHFYPLAGLKENISIDCNDGADFLKRVNPSLSLLPSSLSLOQLI 401
Qy 121 PIKPSMEASIDERNPLIGVQNVDPDSG-IAIGSVSHKLIDGGTADCFKSWGAV-FRG 178
Db 402 PTVSDSIET-----RTRLLAQASFFECGSMGIVCISHKLADATSIGLFMSNAISSRG 457
Qy 179 CRENIILHPSLEALLPDP---RDDLPKYVDQMEALFAGKVKVATRFVGVKAISSIQ 235
Db 458 SIKTIGAP-VDTVKIPFGNPFSTSPAPVPEPEIMM---NQTLSKRFIFDSSISIALQ 512
Qy 236 DEAKSEVPKPSRVHVTGFLWKHLIAASRALTSQTSTRLSIAQAQVNLRTMMETVL 295
Db 513 AKASSFVQNPQTRVEAVSALIKSMKATRV-SGT--SKPSILANSASLSRVS-PPPT 569
Qy 296 DNATGNLFWQAQILIEHSHTTPEISDLKDLVNLINGSVKQNGDYFETPKGEGYGR- 354
Db 569 KNSIGNLVSVFAKAE-----EGINQTKQLTVSKIRKAKQRFROIHIPKLVGNPNATEI 623
Qy 355 MCEYLDQRTWSSMEAP--DIYLFSSWTNF-FNPLDQGWGRTSWIGVAGKIESASCKFI 411
Db 624 ICSTY---QKAGDMIAGDDFYIFSSACRGLVETDFGKGVKPVWVGPPSVRQK---NIV 677
Qy 412 ILVPTQCGSGIEAVNLEERKQWMLQDPHFILALAS 447
Db 678 TLLDTKEAGGIEAVNLEQEMNLFEQDRELLQFAS 713

RESULT 8
AAG27610
ID AAG27610 standard; protein; 823 AA.
XX AC AAG27610;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 32515.
```

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

XX 25-MAR-1999; 99US-0126264P.

XX 29-MAR-1999; 99US-0126785P.

XX 01-APR-1999; 99US-0127462P.

XX 06-APR-1999; 99US-0128234P.

XX 08-APR-1999; 99US-0128714P.

XX 16-APR-1999; 99US-0129845P.

XX 19-APR-1999; 99US-0130077P.

XX 21-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 28-APR-1999; 99US-0130891P.

XX 30-APR-1999; 99US-0131449P.

XX 30-APR-1999; 99US-0132048P.

XX 04-MAY-1999; 99US-0132407P.

XX 05-MAY-1999; 99US-0132484P.

XX 06-MAY-1999; 99US-0132485P.

XX 06-MAY-1999; 99US-0132486P.

XX 07-MAY-1999; 99US-0132487P.

XX 11-MAY-1999; 99US-0132863P.

XX 14-MAY-1999; 99US-0134256P.

XX 14-MAY-1999; 99US-0134218P.

XX 14-MAY-1999; 99US-0134219P.

XX 14-MAY-1999; 99US-0134221P.

XX 18-MAY-1999; 99US-0134370P.

XX 19-MAY-1999; 99US-0134768P.

XX 20-MAY-1999; 99US-0134941P.

XX 21-MAY-1999; 99US-0135124P.

XX 24-MAY-1999; 99US-0135353P.

XX 25-MAY-1999; 99US-0135629P.

XX 27-MAY-1999; 99US-0136021P.

XX 28-MAY-1999; 99US-0136392P.

XX 01-JUN-1999; 99US-0136782P.

XX 03-JUN-1999; 99US-0137222P.

XX 04-JUN-1999; 99US-0137528P.

XX 07-JUN-1999; 99US-0137502P.

XX 08-JUN-1999; 99US-0137724P.

XX 10-JUN-1999; 99US-0138094P.

XX 10-JUN-1999; 99US-0138540P.

XX 10-JUN-1999; 99US-0138847P.

XX 14-JUN-1999; 99US-0139119P.

XX 16-JUN-1999; 99US-0139452P.

XX 17-JUN-1999; 99US-0139453P.

XX 18-JUN-1999; 99US-0139492P.

XX 18-JUN-1999; 99US-0139454P.

XX 18-JUN-1999; 99US-0139455P.

XX 18-JUN-1999; 99US-0139456P.

XX 18-JUN-1999; 99US-0139457P.

XX 18-JUN-1999; 99US-0139458P.

XX 18-JUN-1999; 99US-0139459P.

XX 18-JUN-1999; 99US-0139460P.

XX 18-JUN-1999; 99US-0139461P.

XX 18-JUN-1999; 99US-0139462P.

XX 18-JUN-1999; 99US-0139463P.

XX 21-JUN-1999; 99US-0139750P.

XX 21-JUN-1999; 99US-0139763P.

XX 22-JUN-1999; 99US-0139877P.

XX 22-JUN-1999; 99US-0139899P.

```
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140696P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142300P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145818P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.

PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161363P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161923P.
PR 28-OCT-1999; 99US-0161924P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.5%; Score 533.5; DB 3; Length 823;
Best Local Similarity 33.3%; Pred. No. 1.5e-45;
Matches 152; Conservative 87; Mismatches 180; Indels 37; Gaps 18;

QY 3 KIEVINSKHTIKPSTSTP--LQPYKLTLLDQITPPAYVPVIVFFYPTIDHDFNLPTOLA 60
Db 381 QMRVDVVSRIIKPS-SPTENHLKKFKLSLEQLGPTIFGPMVFFYS-ANNSIKPTSLQ 438
QY 61 DLROALSETLTYPLSGRVNNLYIDDPFEGVPYLEARVNCMDTDFIRLKRKIECLNEFV 120
Db 439 MLKKSLSETLTHEYPLAGRLKGNISIDCNDSGADFLFARVNSPLSLLEPPSSDSLOQLI 498
QY 121 PIKPFMEALSDERYPILGVQVNVDSG-TAIGVSVSHKLIDGTDACELKSMGAV-FRG 178
Db 499 PTVDSIET----RTRLLLAQASFFECGMSIGVCSHKLADATSIGLFKSWAAISRG 554
QY 179 CRENIIEPSLSEAAALLPPP---RDDLPEKYVDQMEALWPAKKVATRFVFGVKAISIQ 235
Db 555 SIKTIGAP-VFDTVKIAPPNGFNSETSPAPVVEIYK---NQTLKRFIFDSSSIQALQ 609
QY 236 DEAKSESVKPSRVHVTGFLWKHLIAASRALTSGTSTRLSTAAQAVNLRTNMNMTVL 295
Db 610 AKASSPEVNQPTREAVSALIKWSAMKATRV-SGT--SKPSTIANASLSRVS-PPFT 665
QY 296 DNATGNLFWQAILELSHTTPEISLKLCDLVNLLNGSVKQCNGDYFETFKGEGYGR- 354
```

```
Db 666 KNSIGNLVSYFAKAE-----EGINTKLTQTVSKIRKAKQRFROIHIPKLVGNPNATRI 720
QY 355 MCEYLDFOQRTSMSEBAP--DIYLFSSWTFN-FNPLDFQWGRTSWIGVAGKTESASCKXI 411
Db 721 ICSY---QKEAGDMIASGDFDFYIFSSACRFGLYETDFQWGPVWVGFFSVROK---NIV 774
QY 412 ILVPTQCGSGIEAWNLEEKAMLEQDPHFLALAS 447
Db 775 TLLDTKAGIEAWNLEQNMNLEQDRELLOQFAS 810

RESULT 9
AAG27609
ID AAG27609 standard; protein; 954 AA.
XX AC AAG27609;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 32514.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-01231180P.
XX PR 09-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126284P.
XX PR 29-MAR-1999; 99US-0126785P.
XX PR 06-APR-1999; 99US-0127462P.
XX PR 08-APR-1999; 99US-0128234P.
XX PR 16-APR-1999; 99US-0128714P.
XX PR 19-APR-1999; 99US-0129845P.
XX PR 21-APR-1999; 99US-0130077P.
XX PR 23-APR-1999; 99US-0130449P.
XX PR 28-APR-1999; 99US-0130891P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 30-APR-1999; 99US-0132407P.
XX PR 04-MAY-1999; 99US-0132484P.
XX PR 05-MAY-1999; 99US-0132485P.
XX PR 06-MAY-1999; 99US-0132486P.
XX PR 07-MAY-1999; 99US-0132487P.
XX PR 11-MAY-1999; 99US-0132863P.
XX PR 14-MAY-1999; 99US-0134256P.
XX PR 14-MAY-1999; 99US-0134218P.
XX PR 14-MAY-1999; 99US-0134219P.
XX PR 14-MAY-1999; 99US-0134370P.
XX PR 19-MAY-1999; 99US-0134941P.
XX PR 20-MAY-1999; 99US-0135124P.
XX PR 21-MAY-1999; 99US-0135353P.
XX PR 24-MAY-1999; 99US-0135629P.
XX PR 25-MAY-1999; 99US-0136021P.
XX PR 27-MAY-1999; 99US-0136332P.
XX PR 28-MAY-1999; 99US-0136782P.
XX PR 01-JUN-1999; 99US-0137222P.
XX PR 03-JUN-1999; 99US-0137528P.
XX PR 04-JUN-1999; 99US-0137502P.
XX PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0142977P.
PR 14-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144864P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145152P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
```



activity for producing and regulating aromatic and/or aliphatic ester formation in microorganisms, plant cells or plants.

Example 5; Page 116; 163pp; English.

The present invention describes nucleotide sequences with thiolase, alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase, aminotransferase and esterase activities, which are involved in the biosynthetic pathway for aliphatic and/or aromatic ester production in fruit. The nucleotide sequences can be inserted into the genome of a fruit-producing plant to regulate aliphatic and/or aromatic ester formation. Aromatic and/or aliphatic esters in microorganisms, plant cells or plants are produced by inserting thiolase, alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase, aminotransferase and esterase nucleotide sequences into the genome and feeding the microorganism or plant with alcohol, aldehydes, alpha-keto acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and their proteins can be used in the processed food industry as food additives to enhance the flavour of syrups, ice-creams, frozen desserts, yoghurts and confectionery. They are used: as flavouring agents for oral medications and vitamins; provide flavour and aroma in beverages, including alcohol; enhance or reduce fruit flavour, aroma, fragrance or scent; enhance the flavour or aroma of natural, synthetic or artificial products; for the production of novel combinations of artificial flavour substances; as antibacterial or anti-fungal agents; as fragrance or perfumes in cosmetics, creams, sun-protectant products; hair conditioners, lengthening agents and fixatives in perfumes, suspension aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning products, personal care products and animal care products; as disinfectant additives; as degreasing solvents for electronics; as insect pheromones; and as dye carriers, solvents, insect repellents, miticides, scabicides, plasticisers and decolorants. The present sequence represents a novel alcohol acyl transferase, from the present invention.

Sequence 426 AA;

[illegible]

RESULT 11  
ABB92488

ABB92488  
ID ABB92488 standard; protein; 443 AA.

AC ABB92438;

DT 31-MAY-2002 (First entry)

DE Herbicidally active polypeptide SEQ ID NO 1699.

KW Herbicidal; plant; agriculture; herbicide.

CS Arabidopsis thaliana.

XX PN WO200210210-A2

XX  
PD 07-FEB-2002

PF 28-AUG-2001: 2001WO-EP009892.

XX  
PR 28-AUG-2001: 2001WQ-EP009892

XX PA (FARB ) BAYER AG.

XX PI Tietjen K. Weidler M:

XX  
DR WPT: 2002-269070/37

XX  
PT  
Identifying plant fa

PT comprising aligning acid comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms.

PS Claim 5; SEQ ID NO 1699; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins (ABB90790-ABB94016; for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected. The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides

Sequence 443 AA;

Query Match 21.0%; Score 498.5; DB 5; Length 443;

Matches	152;	Conservative	89;	Mismatches	171;	Indels	71;	Gaps	23;
Best Local Similarity					31.5%;	Pred. NO. 2.4e-42;			

QY 1 MEKIEVSINSKHTIKPSTSSTP--LOPYKLTLLDCLTPPAYVPIVEFYPIITRHDENI.P-Q 57

Db 1 METMKVETIAKEI IKPS-STTPNDLOTLOLSIYTHILPPVYTVAEJFY--TKDDIJSPEC 57

QY 58 TLADLRQALSETLTLYPLSGRVKNNLYIDDFEEGWPYLEARV-NCDMTDFLRPKTECI. 116

db 58 SSKIKTSLART.TKFPYI.AGRK-GVTIDCNREGAVENDAPUNNYPI.SDR:PSDPDEKTI.116

117 NEED THE PRESIDENT TO BE CONVINCED - GOVERNMENT TO DO IT. DO NOT WANT TO

[illegible]

176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 115

[illegible][illegible][illegible][illegible][illegible]

Db 287 LLAESLIGNIM-FSSVILGIS-----QOEIKIEKAVRDLR---KQ--GDDLQNVIKBEQG 336  
 Qy 352 -----YGRMCEYLDLDFORTWSSMEPAPDIYLFSSWTNFFNPL----DFGW 391  
 Db 337 SSSAMIGSKLANLMTNYSKL-----SYETHEP-----YTVSSWCKL--PLYEASFGW 382  
 Qy 392 GRTSWIGVAGKIESASCKFIILVPTQCGSGIEAWNVLEEKMMLEQDPHFLALAS--PK 449  
 Db 383 GSPVWI--AGNVAPMLENVMTLIDSKDQGIIEAFVTLPEENMLSLQNTTELLAFASVNP 440  
 Qy 450 TLI 452  
 Db 441 VLV 443

RESULT 12  
 ABB91119  
 ID ABB91119 standard; protein; 436 AA.  
 XX AC ABB911119;  
 XX DT 31-MAY-2002 (first entry)  
 XX DE Herbicidally active polypeptide SEQ ID NO 330.  
 XX KW Herbicidal; plant; agriculture; herbicide.  
 XX OS Arabidopsis thaliana.  
 XX PN NC020210210-A2.  
 XX PD 07-FEB-2002.  
 XX PE 28-AUG-2001; 2001WO-EP039892.  
 XX PR 28-AUG-2001; 2001WO-EP009892.  
 XX PA (PARB + BAYER AG.  
 XX PI Tietjen K, Weidler M;  
 XX DR WPI; 2002-269010/31.  
 XX PT Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms.  
 XX PS Claim 5; SEQ ID NO 330; 261pp + Sequence Listing; English.  
 XX CC The invention relates to identifying target proteins (ABB90790-ABB94016)  
 CC for herbicidally active compounds, comprising aligning and comparing  
 CC nucleic acid or amino acid sequences from plant with nucleic acid or  
 CC amino acid sequences from non-plant organisms using suitable search  
 CC parameters, where plant sequences having an E-value greater by a factor  
 CC of 3 than the E-value of most similar non-plant sequences are selected.  
 CC The polypeptides or nucleic acids encoding them are useful for  
 CC identifying modulators. The identified modulators are useful as  
 CC herbicides  
 XX SQ Sequence 436 AA;

Query Match 21.0%; Score 498; DB 5; Length 436;  
 Best Local Similarity 30.9%; Pred. No. 2.7e-42;  
 Matches 145; Conservative 90; Mismatches 184; Indels 50; Gaps 19;

Qy 1 MEKIEVSINSKHRTKESTSTP--LQPYKLTLDLQDTPPAYVPIVFPYPTDHD---FNL 55  
 Db 1 MEK-NVEILSREIVKPS-SPTDDKRLNLSLLDLSSEPMYTGALLFYADPQNLGFSF 58  
 Qy 56 PQTLLADLQALSETLTYLPLSRGVKNLVLDIFEGVPLVLEARNVCDMTDFLRLEKIEC 115  
 Db 59 EETSLSLKKLSKLTPIFYPLAGRIIGS-FVECNDEGAVFEIARVDHLLSEFLKCPVPS 117

Qy 116 LNEFVPIKPSMEALSDERYPLLGHQWVFD-SGIAIGVSVSHKLDGTDACFLKSWGA 174  
 Db 118 LELLIPVEAKREAVT---WPVLLIQANFFSCGGLVITICVSHKITDATSLAMFIRGNAE 174  
 Qy 175 VFRGCRENIIPSLSEAAALLPPRDCLPEKYVDOMEALWFAGKKVATREFFVGVKAISII 234  
 Db 175 SSRGIGITLI-PSFTASEVFPKPLDELSPKMDRAEEV--EEMSCVTKEFVFDASKIKKL 231  
 Qy 235 QDEAKESVPKPSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAAQAVNLTRMMETV 294  
 Db 232 RAKASRNLVKPNTRVEAVTALFWRCVTKVSL---SSLTPRTSVLQILVNLGRK--VDSL 286  
 Qy 295 LDNATGNLFWN-----AQALELSHTTPEISDLKCLDLVNLNGSVKQCNDYFETPK- 347  
 Db 287 CENTIGNMLSLMILKNEEAARIQDVVDEIRRAK--BIFSLNCKEMSKSSRIPELLEE 344  
 Qy 348 -GKEGYGRMCEYLDLDFORTWSSMEPAPDIYLFSSWTNF--FNPLDFCWGRTSITGVAGKIES 405  
 Db 345 IGKV-YRGENEM-----DLWMSNCKLGLYDADFQWGPVW--VTGRGTS 387  
 Qy 406 ASCKFIILVPTQCGSGIEAWNVLEEKMMLEQDPHFLALAS--PKTLI 452  
 Db 388 HFKNMLMLIDTKDGEIEAWITLTERQMSLFECQDELLESASLNPVLI 436

RESULT 13  
 ABB91973  
 ID ABB91973 standard; protein; 436 AA.  
 XX AC ABB91973;  
 XX DT 04-DEC-2003 (first entry)  
 XX DE Acyltransacylase enzyme #11.  
 XX KW Acyltransacylase; transacylase; paclitaxel; taxoid;  
 KW paclitaxel biosynthetic pathway; enzyme.  
 XX OS Arabidopsis thaliana.  
 XX PN US2003108891-A1.  
 XX PD 12-JUN-2003.  
 XX PR 18-SEP-2002; 2002US-00166984.  
 XX PR 30-SEP-1999; 99US-00411145.  
 XX PR 07-DEC-1999; 99US-00457046.  
 XX PR 25-MAY-2001; 2001US-00866570.  
 XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
 XX PI Croteau RB, Walker KD, Schoendorf A, Wildung WR;  
 XX WPI; 2003-659103/62.  
 XX PT New purified protein useful for producing paclitaxel and related taxoids.  
 XX PS Example 1; Fig 6; 135pp; English.  
 XX CC The invention relates to a transacylase enzyme and the polynucleotide  
 CC encoding it. The invention also relates to a recombinant nucleic acid  
 CC molecule comprising a promoter sequence operably linked to the  
 CC polynucleotide encoding the transacylase enzyme and a cell transformed  
 CC with the recombinant nucleic acid. The transacylase enzyme is useful for  
 CC producing paclitaxel, related taxoids and intermediates in the paclitaxel  
 CC biosynthetic pathway. This sequence represents an acyltransacylase enzyme  
 CC of the invention.  
 XX SQ Sequence 436 AA;

Query Match 21.0%; Score 498; DB 7; Length 436;





Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	498	21.0	436	3	US-09-457-046B-62	Sequence 62, Appl
2	485	20.4	446	3	US-09-457-046B-70	Sequence 64, Appl
3	448	18.9	435	3	US-09-457-046B-66	Sequence 60, Appl
4	404.5	17.0	433	3	US-09-457-046B-66	Sequence 66, Appl
5	387.5	16.3	439	3	US-09-457-046B-68	Sequence 68, Appl
6	308	13.0	331	3	US-09-457-046B-59	Sequence 59, Appl
7	242	10.2	461	3	US-09-457-046B-64	Sequence 64, Appl
8	230	9.7	448	1	US-08-207-904-2	Sequence 2, Appl
9	230	9.7	448	1	US-08-207-904-17	Sequence 17, Appl
10	228.5	9.6	451	3	US-09-457-046B-69	Sequence 69, Appl
11	225	9.4	445	3	US-09-457-046B-73	Sequence 73, Appl
12	222	9.5	497	3	US-09-457-046B-65	Sequence 65, Appl
13	219	9.2	482	3	US-09-457-046B-63	Sequence 63, Appl
14	179	7.5	443	3	US-09-457-046B-50	Sequence 50, Appl
15	178.5	7.5	461	3	US-09-457-046B-70	Sequence 70, Appl
16	177.5	7.5	441	3	US-09-457-046B-54	Sequence 54, Appl
17	177.5	7.5	458	3	US-09-457-046B-61	Sequence 61, Appl
18	176.5	7.4	306	3	US-09-457-046B-22	Sequence 22, Appl
19	175	7.4	455	3	US-09-457-046B-72	Sequence 72, Appl
20	173.5	7.3	460	3	US-09-457-046B-71	Sequence 71, Appl
21	157	6.6	438	3	US-09-457-046B-58	Sequence 58, Appl
22	154.5	6.5	450	3	US-09-457-046B-67	Sequence 67, Appl
23	150	6.3	303	3	US-09-457-046B-20	Sequence 20, Appl
24	148.5	6.3	440	3	US-09-457-046B-45	Sequence 45, Appl
25	146.5	6.2	459	4	US-09-538-414-2	Sequence 2, Appl
26	146.5	6.2	459	4	US-10-074-279-2	Sequence 2, Appl
27	144	6.1	439	3	US-09-457-046B-28	Sequence 28, Appl



```

Db 58 LESSLSTLLYVAVGRYKTCVIECNDGIGYVETAFVHLQFLLGBESNHLILV 117
QY 121 PIKFSMEALSDERYPLLGQVNVFD-SGIAIGVSVSHKLDGTDACFLKSWGAVRGC 179
Db 118 GLSGFLSET--ETPLAAIQLNFKCGGLVIGAFQFNHIIIGDMFTSMNSWAKACKVG 174
QY 180 RENIHLSLSEAILL-----FPRDDLPEKYDQMEALWFAGKAVATRRFVGVKALS 232
Db 175 IKEVAHTFTGLAPLPSAKVINIPPPS-----FEGVKFVKRKFVFNENAIT 221
QY 233 SIQDEAKSE-----SVKPSRVHVTGLMKHLIAASALTSGLTSTLSIAAQA 284
Db 222 RLKKEATEEDGDGDDQKKRPSVDLVTAFLSKSLIEMDCAKKEQTKS-RPSLWVHMN 280
QY 285 LRTRMNMETVLDN-ATGNLFWAAQILLESHTTPEISDL-----KLCD- 326
Db 281 LRKRTKL--ALENDVSGNFFTVNAESKIT-VAPKITDLSLGSACGEIISVAKYDDA 337
QY 327 --LVNLLNGSVKQCNQGYFFTFKCKEGGRMCEYLDLQRTWSSMEPAPDIYLFSSWTF 384
Db 338 EVVSSWINSVREF--YYEKGKGEK-----NPLITSWCKRF- 371
QY 385 NPL---DFGWRGRTSWIGVAGKIESASCKF--IILVPTQCGSGIEAWNLBEKAMLEQ 438
Db 372 -PLYEVDSGM-----GIPSLVDITAVPFGLIIVLMDERAPAGDIAVRACTSEHDMIQPQ 424
QY 433 DPHFLALAS 447
Db 425 HHQLLSYVS 433

```

RESULT 5  
 US-09-457-046B-68  
 ; Sequence 68, Application US/09457046B  
 ; Patent No. 6287835  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Croteau, Rodney et al.  
 ; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
 ; FILE REFERENCE: 53679  
 ; CURRENT APPLICATION NUMBER: US/09/457,046B  
 ; CURRENT FILING DATE: 1999-12-07  
 ; NUMBER OF SEQ ID NOS: 74  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 68  
 ; LENGTH: 439  
 ; TYPE: PRT  
 ; ORGANISM: Catharanthus roseus  
 US-09-457-046B-68

```

Query Match 16.3%; Score 387.5; DB 3; Length 439;
Best Local Similarity 26.8%; Pred. No. 3e-33;
Matches 128; Conservative 83; Mismatches 196; Indels 71; Gaps 21;

QY 1 MEKIEVSIN---SKHTIKPSTSTSP--LQPKLTLQLDLPAPXVPIVFFVPIPTHDFN 54
Db 1 MESKISIVETLTSLKILKPS-SPTQSLSRYNLSYNQNIYQTCVSGVFFVFNPD-GIE 58
QY 55 LPQTLADLRQALSETLLTYFLSGRVKXNLYIDFEFEGVPYLEARVNCMDTDFLRKIE 114
Db 59 ISTIREQLQNSLSTVSYYPFAGKVKNKYIHQNDNDGIEFVEVIRCRMDILKYE--- 115
QY 115 CLNBEVPIKPSMEALSDERYPLLG-----VQVNVFD-SGIAIGVSVSHKLDGTDAC 167
Db 116 -----LRSYARQVLVLPKRVTVGSEDTTAVQLSHFDGGLAVAFG:SHKVADGGTAS 168
QY 168 FLKSW--GAVRGGRENIIHPSLSAALLFPPRDLLPEKYVDQMEALWFAGKAVATRRFV 225
Db 169 FMKWAASACVLSSSHVPTLLVSDS--IFPQDNII---ICEQFP-----TSKNCVEKFI 220
QY 226 FGVKAISIQDEAKSESVPKSRVHVTGLMK-HLIAASALTSGLTSTLSLSIAAQA 283
Db 221 FPEAIEKLKSKAVFEGIEKPTRVEVLNFTLSRCATVAGKSAKNNCQSLPFFVLQAI 280

```

```

QY 284 NLRTMNMETVLDNATGNLFWAAQILLESHTTPEISDLKCLDLVNLNGSVKQCNQGYD 343
Db 281 NLRTILP---QNSVGNL-----VSIYFSRTIKENDYLNKEYTKLVINELRKEKQIK 332
QY 344 ETFKKEGIGRMCYLDLQRTWSSMEPAP---DIYLFSSWTFN-FNPLDFGWRGRTSWIG 398
Db 333 NLSREKLYTYAQM--EFVKSLEFDSLNFLLDAYLSDSWCRFPFYDVFDFGKGPVW- 389
QY 399 VAGKIESASCKP-----IILVPTQCGS--GIEAWNLBEKAMLEQDHFLLALAS 447
Db 390 -----CLFQPIKXNVVMYDPFGDDYGIEALVSEFQEKMSAFKXNQLLQFVS 438

```

RESULT 6  
 US-09-457-046B-59  
 ; Sequence 59, Application US/09457046B  
 ; Patent No. 6287835  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Croteau, Rodney et al.  
 ; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
 ; FILE REFERENCE: 53679  
 ; CURRENT APPLICATION NUMBER: US/09/457,046B  
 ; CURRENT FILING DATE: 1999-12-07  
 ; NUMBER OF SEQ ID NOS: 74  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 59  
 ; LENGTH: 331  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-457-046B-59

```

Query Match 13.0%; Score 308; DB 3; Length 331;
Best Local Similarity 26.8%; Pred. No. 8.5e-25;
Matches 93; Conservative 70; Mismatches 144; Indels 40; Gaps 15;

QY 117 NEFVPIKPSMEALSDERYPLLGQVNVFD-SGIAIGVSVSHKLDGTDACFLKSWGAV 175
Db 11 NELNKLHPFEHVSDD--VPLT-VQLTFEGGIALGIGLSHKLCTALSGLIFVNSWAF 67
QY 176 FRGCRENIHPSLSAALLFPPRDLLPEKYVDQME-ALWPAKGVATRRFVGVKALSSI 234
Db 63 ARGQTDLEIITPSF-DLAKMFPFCD-----IENLNMATGITKENIVTRRFVFLRSVESL 120
QY 235 QDEAKSESVPKSRVHVTGLMKHLIAASALTSGLTSTLSIAAQAANLRTMNMETV 294
Db 121 RERPSGNKKIRATRVESLVSVFWSRFMAST---NHDDKTGKIYTLIHPVNLRRQADD-I 176
QY 295 LDNATGNLFWAAQILLESHTTPEI---SDLKCLDLVNLNGSVKQCNQGYFETFKGKE 350
Db 177 PDNMFEN-----IMRFSVTVPMXINENDEKASLVDDQMEERKIDAVYVKLIQ-RJ 228
QY 351 GYGRMCEYLDLQRTWSSMEPAPDIYLFSSWTFNPL---DFGWRGRTSWIGVAGKIESAS 407
Db 229 NRG-----HLEFLNKQASGFVNGEIVSFTSLCKFPVVEADFGWKPLWVASA---RMSY 281
QY 408 CKFIILVPTQCGSGIEAWNLBEKAMLEQDHFLLALAS--PKTIL 452
Db 282 KNLVAFIDTKEGDGIEAWNLQNDMSRFEADELLRYVSSNPSVWV 328

```

```

RESULT 7
US-09-457-046B-64
; Sequence 64, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1

```



```

; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-69

Query Match      9.6%   Score 228.5; DB 3; Length 451;
Best Local Similarity 22.6%; Pred. No. 6.4e-16;
Matches 99; Conservative 70; Mismatches 162; Indels 87; Gaps 19;

Qy 13 TIKPSTSSPTLPQ-----YKLTLDQLTPPAYVPIVFYFYPITDHDNLPQTLLADLRQALS 67
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 5 TPKSXTIVPAEFTMSGRFFLAEMWQQGTTHIPTLYFYDKPSESFO-GNVVEILKTLSL 63
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 68 ETTLTYLYPLSGRV-----KNNLNYIDDPEEGVPYLEARVNCMDWDLRLRKIECLNEFEVPIK 123
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 64 RVLVHFPYAGRLRNLPGRPFELNCNAEGVEFIAESEGLSDP-----KDFSPPT 114

Qy 124 PFS--MEAIS-----DBRYPLLGGVQNVFD-SGIAIGVSVSHKLIDGTADCFKLSWGAVF 176
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 115 ERENLMPOVNKYKNPIETIPLFLAQVTKFCKGGISLVNVSHAIVDGQSALHLISEWGRLA 174

Qy 177 RGCENIIHPSLEAAL-----LFPR-----DDLPEKYVDOMEALWFACKKVATR 222
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 175 RGSEPLETW-PFLDRKXILWAGEPLFPFVSPKPDHFEDQPPLFGTNDVBERKKKTIIV 233

Qy 223 RFVFGVKAISSIODEAKSESVPKP----SVHAVTGFLWKHLIAASRALTSGETTSLRSI 278
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 234 MLFLSTSOLQKLRSKANGSKHSDFAPKFTYETVGVHWVR--CAKARGHSPFOPTALGI 291

Qy 279 AAQAVNLRTMMNETVLDNATGMFLFWAAQAILEL--SHTTPEISDLKLCDAWN:LINGSVK 336
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 292 C-----IDTRSEVPPLPRG-----YFNATLDWVAASTSGELISNELGFAASLISAIAK 341

Qy 337 QCGNDYFETPKHGKGRCMCEYLDFQATSMSEPAPDIYLFSSWTNFF--NP-----386
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 342 NVTNEYVMI-----GIEYLKNQDXLKKFO---DLHALGSTGEPPGYGNPN:GVVSWL 389

Qy 387 -----LDFCWGRTSWG 398
    ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 390 TLPMYGLDLPGWGREFTYG 407

RESULT 11
US-09-457-046B-73
; Sequence 73, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Pacitaxel Biosynthetic Pathway
; FILE REFERENCE: 33679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 73
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-73

Query Match      9.5%   Score 225; DB 3; Length 445;
Best Local Similarity 22.7%; Pred. No. 1.5e-15;
Matches 106; Conservative 86; Mismatches 181; Indels 94; Gaps 25

Qy 4 IEVSINSEKHNIKSTSTSTPLQPVKLTLLDOL--TPPAVPIVVFYFITDH-----51
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MSTQIKQSTWVRPA-EETPNKSLMLSNIDLIRLTFYSHTGAVALYKQPDNNEDNIHPSSS 59

Qy 52 ---DFNLPOTLADRQALSETLTYPLPSGRVQN---LYIDDFEEGVPIYLEARVNCMDT 105
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 60 MYFDANI-----LJEALSKAIVFPYFMAGRKLINGDRYEIDCNAEGALFVEASSSHVLE 113

Qy 106 DFLRLRKIECLNE-FVPIKPFMSMAISDERYPPLLGQVNVFD-SGIAIGVSVSHKLIDGG 163
    ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 114 DFGDFRPNDELHVRVWPTCDYS-KGISS--FPLLMVQLTRFCGGVSGIFQAHHHVCQGM 170  
Qy 164 TADCFELKSWGAVFEG-----CRENIH-----PSLSAALLFPP-RDDLPKYYDQM 209  
Db 171 AHPEFNNSWARIAGKGLPALPEVDRYHLHPRNPQIKYSHSOFEPFVPSLPELLED-- 228  
Qy 210 EALWFAGKQVATR-FVFGVKALSIIOEAK-SBSVPKSRVHAVTGELMKHLIAASRAL 267  
Db 229 -----GKTNKSQTLFILSRQINTLQKLDLSNNTLSTLSTYEVAAHWRS-VSKARGL 281  
Qy 268 TSGTTSTRLSIAQAVMLRTMMETVLDNATGNLFWMAQAILELSHTTPEISDLK--L 324  
Db 282 -SDHEEIKL---IMPVGRGRINPSPKPGYCGN-----WFLAVCTATVGLSCNPL 330  
Qy 325 CDLVNLINGSVKCKGQYFFETFKGEGYGRMCEYLDQRTMSSM-----EPAPDIY--- 375  
Db 331 TDTAGKQVQKALGLDDYLS-----AIDHTSKPGLPVPMYSGPEKTLFNV 378  
Qy 376 LFGSWTNF-FNPLDFGWRGTSWIGVAGKIESASCKFIILVPTQCGSG 421  
Db 379 LVNSWGRIPYQAMDFGWSPTFFGISNIFVDQC---FLIPSRDGDG 422

RESULT 12  
US-09-457-046B-65  
; Sequence 65, Application US/09457046B  
; Patent No. 6287835  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney et al.  
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
; FILE REFERENCE: 53679  
; CURRENT APPLICATION NUMBER: US/09/457.046B  
; CURRENT FILING DATE: 1999-12-07  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 65  
; LENGTH: 497  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-457-046B-65

Query Match 9.4%; Score 222; DB 3; Length 497;  
Best Local Similarity 22.9%; Pred. No. 3.8e-15;  
Matches 115; Conservative 86; Mismatches 159; Indels 142; Gaps 28;  
Qy 6 VSINSKHTIKPSTSTPLQPKYKTL--LDQLTTPPAYVPIVFFYPIITDHDNLPOTLA--D 61  
Db 55 VTTRKQVI---TAALPLQDHWLPLSLNLDLLPLNTHVCECYKKPLFTN---TVAYET 108  
Qy 62 LQALSETLTYLPLSGRVKN-----LYIDDFEGVPIYLEARVNCMDTDFLELRKIEC 115  
Db 109 LKTALETLVSYTAFAGELVTNPTGEPEILCNN--RGVDFVEAGADVLELRNLYDPDES 166  
Qy 116 LNEFVPIKPSMEALSDERYPLLGVOVNVDSG-IAIGSVSHKLIIDGTADCFKLSWGA 174  
Db 167 IAKLVPIK-----KHGVTAIQVTLQKGSIVVGCFTDHRVADAYSMMNELLWAE 216  
Qy 175 VFR-----GC-----RENIHPSLSEAL-----LFPPRDDL-PEKYVDOMEAL 212  
Db 217 ISKSDVPSICVPSRSLNPRRLVMDPSIDQIYMPVSLPPQETTNENL----- 269  
Qy 213 WFAGKQVATRRFVFGVKALSIIOEAKSBSVPKSRVHAVTGFLWKHLIAASRALTSQTT 272  
Db 270 -----LASRTYITKANALQBLQTLASSKNGKRTKLESFSAFLWK-LVAHAARDVPVI 322  
Qy 273 ST-RLSTAAQAVNLRTMMETVLDNATGNLF---WMAQAILE-----LSHTTPEISDLK 323  
Db 323 KTSKLG1---VVDGRRKL-MEKENVTFGNVLSPVFGQRIIDLLISPLSVWTEVH--- 375  
Qy 324 LCDLVNLINGSVKCKGQYFFETFKGEGYGRMCEYLDQRTMSSMEPAPDIY--- 374  
Db 376 -----RLKKSVT-----KEHFLNLIDWVE-----TCRPTPAVSRIYVSGSD 412

Qy 375 -----YLFSSWTFN-FNPLDFCWG-----RTSWIGVAGKIESASCKFIILVPTQCGSGI 422  
Db 413 DGPAFVSSCRSPVNVQDFGWSPEVFGSYHFPGWGSAG-----YVMDMP-----SSV 460  
Qy 423 B-----AWNVLEBEKMALEQD 439  
Db 461 DDDRMVYLHLTKGLRFIEEE 482

RESULT 13  
US-09-457-046B-63  
; Sequence 63, Application US/09457046B  
; Patent No. 6287835  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney et al.  
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
; FILE REFERENCE: 53679  
; CURRENT APPLICATION NUMBER: US/09/457.046B  
; CURRENT FILING DATE: 1999-12-07  
; NUMBER OF SEQ ID NOS: 74  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 63  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-457-046B-63

Query Match 9.2%; Score 219; DB 3; Length 482;  
Best Local Similarity 22.5%; Pred. No. 7.7e-15;  
Matches 122; Conservative 93; Mismatches 171; Indels 156; Gaps 26;  
Qy 1 MEKIEVSINKHTIKPSTSTPLQPKYKTLDDQLTTPPAYVPIVFFYPI-----TDHDFN 54  
Db 4 LEX-SVTIIISNRVFPDQKST-LVDLKLVSVD-----LPMLSCHVYIQKGLCTCPNLP 54  
Qy 55 LPQTLADLQALSETLTYLPLSGRV---KNLNYIDDFEGVPIYLEAR-----V 100  
Db 55 LPALISHLKSLSLITLTHFPPLAGLSTSSGSHVFLTCNDAGDFVFAQAKSHVSDVIA 114  
Qy 101 NCDMTDFLRLRKIECLNEFVPIKPSME-ALSDERY--PLLGVOVNVFDSGIALGVSVSH 157  
Db 115 GIDVPDVVK-----EF-----FTYDRAVSVEGHNRPILAVQVTELDNGVFIGCSVNH 161  
Qy 158 KLIDGGTADCFKLSWGAVFRCGRNIHPSLS-EAALLFPPRDDLPKYYVDOMEALWPAQ 216  
Db 162 AVTDGTSLWNFINPTFAEVSGRGAKNVTRQPDFTRESVLISPAVLKVPQ-----SG 210  
Qy 217 KYVA-----TFRFVFGVKAISIQ-----DE--- 237  
Db 211 PKVTFDENAPLREKIFSRFSRESIOELKAVNVNKKMLTVDNGEIDGVELLGQSNKLMGK 270  
Qy 238 -----AKSESVPKP-----SRVHAVTGFLWKHLIAASRALTSQTTSTRLSIA 279  
Db 271 ENGLITLMSLSLFGNDVAVSKPAVEISSFQSCALLWRAITRARKLPSSKXTTFRM--- 327  
Qy 280 AQAVNLTRNME---TVLDNATGNLFWMAQAILELSHTTPEISDLKLC-DLVNL----- 330  
Db 328 --AVNCEHRLSPKLNPEYFGNAIQSVPTFAAABWLSR-----DLKWCADQLNQSAAH 379  
Qy 331 LINGSVKQNGDYFETFKGEGYGRMCEYLDQRTMSSMEPAPDIYLFSSWTFNFPNPLDFG 390  
Db 380 ODGIRSVIADW-----EANPRCFPLGNADGASVTMGSSPRFPMYDN-----DFG 424  
Qy 391 WGRTSWT--GVAGKIESASCKFIILVPTQCGSG-IEAWNVLEBEKMALEQDPHFJALAS 447  
Db 425 WGRPVAVRSGRSNKFDCKISAF-----FREGNGTVDLVVLSPETMAGIESDGEFMRVYT 480  
Qy 448 PK 449  
Db 481 MK 482

RESULT 14

US-09-457-046B-50  
; Sequence 50, Application US/09457046B  
; Patent No. 6287835

GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

; FILE REFERENCE: 53679

; CURRENT APPLICATION NUMBER: US/09/457,046B

; CURRENT FILING DATE: 1999-12-07

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 50

; LENGTH: 443

; TYPE: PRT

; ORGANISM: Taxus cuspidata

US-09-457-046B-50

Query Match

Best Local Similarity 7.5%; Score 179; DB 3; Length 443;

Matches 101; Conservative 64; Mismatches 176; Indels 84; Gaps 21;

QY 1 MEK-----LEVSEKSHITKPTSTPLQYKLTLLDQTPP-AVPIVFFYPITDHDEN 54

DB 1 MEKSSADLHWNIIRVVVAP-CQTPKTIQLQSSIDKMGGFANVLVFG---ASHGV 56

QY 55 LPQTLD---LRQALSETLTYPLSGRVK---NNLYIDDPEGVPLRVNCDMTD 106

DB 57 -----ADPAKTIRBALSKTIVFVFPAGRLRKEDGIEVEGIEQALFVEMAND--- 108

QY 107 FLRLKIECLNEFVIKPFSEMAIS-DEHYPP---LLGVQVNFV-SGIAIGVSVSHKLD 161

DB 109 ---LSVRDLDEYNLFRQLQSSLDTDYKDLHMTVQVTFPGGVNMTSVHQSICD 165

QY 162 GGTADPCFLKSWCAVPRGCRNIIHPSLSEAALLFPDRDLP-EKYVDQMEALWFAGKVA 220

DB 166 GNLGCFKFKMAEIVRGEVVKPSEIPIWRE--LVKPEDVIHLQLYSEIRPLVVEKVG 223

QY 221 TRRFVFGVKAISIOEAKSESVKPSRVHVTGFLWKHLIAASRALTSQTSTRLSIAA 280

DB 224 QTSLSVSEKINHAKCIEESKESSESPFIVTAMVW---LARTAFQIPHN----- 272

QY 281 QAVNLRTRMMETVLD-----NATGNLFWMAQAILELSTTPEISDLKCLDLNLLNGS- 334

DB 273 EDVTLILAMDARRSPDPIPKGYGNVI-----GTYAKD---NVNHLUGSL 317

QY 335 -----VKQCGDYPTFKGEGYGRMCEYLDQFRTMSSMEPAPDIYLFSSWTFN-FNPL 387

DB 318 LHALTVIKSMSSFYENMTSR-----VLVNFSTLDLSMKYENVVLSLSDWSRLGHNEV 369

QY 388 DFGWG 392

DB 370 DFGWG 374

RESULT 15

US-09-457-046B-70

; Sequence 70, Application US/09457046B

; Patent No. 6287835

GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

; FILE REFERENCE: 53679

; CURRENT APPLICATION NUMBER: US/09/457,046B

; CURRENT FILING DATE: 1999-12-07

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 70

; LENGTH: 461

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-457-046B-70

Query Match

7.5%; Score 178.5; DB 3; Length 461;

Best Local Similarity 24.3%; Pred No. 1.8e-10;  
Matches 118; Conservative 68; Mismatches 197; Indels 103; Gaps 26;

QY 1 MEKIEVSINSKHTIKPSTSTPLQYKLTLLDQTPP-AVPIVFFYPITDHDENLQTLA 60

DB 13 LEKKEVEL-----VKPS-KHTECHETLSLSLTDN-DPFNEVMYATIVVFVANGKLNDDPVS 65

QY 61 DLROALSETLTYPLSGRV-----KNLYIDDPEGVPLRVNCDMTDFLRKIKIC 115

DB 55 ELRKALSELVHYPLSGKLMSESGKQLQVYLGEVGFVETATSLDSS---LNYIEN 122

QY 116 LNEFVPIK---PFSMEAISDERYPLLGQVNVFD-SGIAIGVSVSHKLDG-GTADCFLK 170

DB 123 LDDVALRLVPEIBIDYESNVCHPLALQVTFACGGFTIGTALTATHAVCDGIGA--- 177

QY 171 SWGAVERGCRNIIHPSLSEAAL-----LFPDRDLPEK-----YVDQMEA 211

DB 178 -----QIIH-ALTELAAGKTEPSVKSWQORERLVGKIDNKPQKVPQSHIDGFLA 225

QY 212 --LWFAGKKVATRRFVFGVKAISSICD-----EAKSESVKPSRVHVTGFLWKHLIAA 263

DB 226 TSAYLPTTDVVTETINIRAGDIXRLKDSMMKECEYLKESF---TTEVLSSYIWK---LR 279

QY 264 SRALTSQTSTRLSIAAQAIVNLRTRMMETVLD-----NATGNLFWMAQAILELSTTPE 318

DB 283 SRALKLNPDG--ITVLGVAVGIR-----HVLDPPLPKGYGNAY-IDVYVEL--TVRE 327

QY 319 ISDLKCLDLNLLNGSVKQC--NGDYPTFKGEGYGRMCEYLDQFRTMSSMEPAPD-IY 375

DB 328 LBESSISNIANEWKKAKTAYBKGYIEELKNT-----RLMRDSDSMFEGVSDGLF 378

QY 376 LFSSWTN--FPNPLDQNGRTSWIGVAGKIESACKFIILVPTQCGSGTBAWYN--LEEE 431

DB 379 FUTDWHNIGWFGSMDPGWNEPVLRELTORESTVHVGMILKPSKSDPSMEGGVAVIKLP 438

QY 432 KWAMLE 437

DB 439 RDAMVE 444

Search completed: July 6, 2004, 13:40:15

Job time : 20.1722 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:38:51 ; Search time 49.964 Seconds

(without alignments)

2816.033 Million cell updates/sec

Title: US-09-857-518A-6

Perfect score: 2373

Sequence: 1 MEKIEVINSKHTKPTSS.....MAMLEQDEHFLASPKTLI 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498	21.0	436	9	US-09-866-572A-62
2	498	21.0	436	9	US-09-866-570A-62
3	498	21.0	436	14	US-10-167-603C-9
4	498	21.0	436	15	US-10-166-984-66
5	485	20.4	446	9	US-09-866-572A-68
6	485	20.4	446	9	US-09-866-570A-68
7	485	20.4	446	14	US-10-166-984-68
8	485	20.4	446	15	US-10-166-984-74
9	448	18.9	435	9	US-09-866-572A-60
10	448	18.9	435	9	US-09-866-570A-60
11	448	18.9	435	14	US-10-166-984-75
12	448	18.9	435	15	US-10-167-603C-7
13	438	18.5	474	12	US-10-167-603C-14
14	438	18.5	474	12	US-10-167-603C-14
15	438	18.5	474	12	US-10-167-603C-16

16	404.5	17.0	433	9	US-09-866-572A-66
17	404.5	17.0	433	9	US-09-866-570A-66
18	404.5	17.0	433	12	US-10-167-603C-9
19	404.5	17.0	433	14	US-10-166-984-66
20	404.5	17.0	433	15	US-10-166-984-66
21	404.5	17.0	433	15	US-10-166-984-66
22	387.5	16.3	439	9	US-09-866-572A-68
23	387.5	16.3	439	9	US-09-866-570A-68
24	387.5	16.3	439	12	US-10-167-603C-8
25	387.5	16.3	439	14	US-10-166-984-68
26	387.5	16.3	439	15	US-10-166-984-68
27	386	16.3	457	14	US-10-259-549-2
28	386	16.3	460	14	US-10-259-549-7
29	385	16.2	408	14	US-10-259-549-15
30	377	15.9	406	14	US-10-259-549-17
31	374	15.8	403	14	US-10-259-549-23
32	357.5	15.1	393	14	US-10-259-549-25
33	308	13.0	331	9	US-09-866-572A-59
34	308	13.0	331	9	US-09-866-570A-59
35	308	13.0	331	14	US-10-166-984-61
36	308	13.0	331	15	US-10-166-984-61
37	307	12.9	251	12	US-10-424-599-152518
38	266.5	11.2	152	12	US-10-424-599-257533
39	259.5	10.9	449	16	US-10-437-963-108868
40	253	10.7	448	12	US-10-424-599-222595
41	253	10.7	449	12	US-10-425-114-56357
42	249.5	10.5	446	16	US-10-437-963-178429
43	249	10.5	467	12	US-10-424-599-237005
44	247	10.4	468	12	US-10-425-114-42259
45	247	10.4	475	12	US-10-425-114-47544

ALIGNMENTS

RESULT 1

US-09-866-572A-62

; Sequence 62, Application US/09866572A

; Patent No. US20020138859A1

; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

; FILE REFERENCE: 53679

; CURRENT APPLICATION NUMBER: US/09/866,572A

; PRIOR FILING DATE: 2001-05-25

; PRIOR APPLICATION NUMBER: 09/457,046

; PRIOR FILING DATE: 1999-12-07

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 62

; LENGTH: 436

; TYPE: FRT

; ORGANISM: Arabidopsis thaliana

US-09-866-572A-62

Query Match 21.0%; Score 498; DB 9; Length 436;

Best Local Similarity 30.9%; Pred. No. 8.9e-44;

Matches 145; Conservative 90; Mismatches 184; Indels 50; Gaps 19;

1 MEKIEVINSKHTKPTSSSTP--LQPKYLLDQDTPPAVPIVFFPITDHD---FNL 55

1 MEK-NVELSREIVKPS-SPTFDDKRIINLSLIDILSSPMYTGALLFYAADPQNLLGFST 58

56 PQTADLADLQALSETLTYPLSGRVKNLYIDDEEGVPYLEARNVCMWDFLRRLKTEC 115

59 EETSLLKLSKSLKTPFYPLAGRIIGS-FVECDNDEGAVFIARVDFHLLSEFLKCPVPS 117

116 LNEFFYKPFPMSEATSDERYPLLVQVNVFD-SGTAIGVSVSHKLDIGTADCFILKSWGA 174

118 LELLIPVEAKSREAVT---WPVLLIQANFFSCGLVITICVSHKITDALSAMFIRGWA 174

175 VFRGCRENIHPSLSGAAALLFPREDLPKPYVDQMEALWFAGKVKVATRFVFGKAISSI 234

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

Sequence 66, Appl

TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-166-984-75

Query Match 18.9%; Score 448; DB 14; Length 435;  
Best Local Similarity 29.8%; Pred. No. 2e-38;  
Matches 137; Conservative 79; Mismatches 193; Indels 50; Gaps 15;

3 KIEVINSKHITKSTSTPIQPKYKLTLLDQTPPAYVPIVFFPIHDDFNLPQTLADL 62  
4 KLEVT--GKEVIKPA-SPSPDRQLQSLDLYCPIYVSTIFFYDLITSESEV--FSENL 58

63 ROALSETLTYPLSGRVKNNLYIDDFEEGVPYLEARVNCMDTDFLRKRLKIECLNEFVPI 122  
59 KLSLSETLSRFYPLAGRIE-GLSISCNDEGAVFTEARTDILLPDLFRLNLTDSLSGFLPT 117

123 KPFSMEALSDEYPLLGQVNVF--DSGIAIGVSVSHKLIIDGGTADCFKLGWAVFRGCR 180  
118 ---LAAGESPAWPLLSVKYTFFGSGGVAVSVSHKICDIAASLVTFVKWATTTAKGK 174

181 EN-----IHPSLSEALLPFRD---DLPEKYVDQMEALWFAKGVATRRFVGV 228  
175 SNSTIEFAATTIYPPPHMVEQFPSTDSNITSKYV-----LKRFPVFP 220

229 KAISIOEAKSESVPKSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAAQAVALRTR 288  
221 SKIAELKHKAASESVVPVTRVEAINSLIWR---CARNSRSNLLIPROAVMWAQMDIRLR 277

289 MNMTVLDNATGNFWAQAILELSHTTPEISDLKCDLVNLLSGVKOCNGDYFTFKG 348  
278 IPSSVAPKDVIGNL---QSGFSLKDAE--SEFEIPEIVATPRKNKRVNEMIKESLQ 331

349 KEGYGRMCYLDPORTMSMEPADYILFSSWTNF-FNPLDFGWTGTSWIGVAGKIESAS 407  
332 NT-IGQSLLSLMAETVSESTE--IDRYIMSSWCRKPFYEVDFGSGSPVWVGYSHTIYN 388

408 CKFIILVPTCCSGGIRAWNLBEKMWALEQDPHFLALA 446  
389 MVGVLLDSKGDGVAEWISLPEEDMSVFVDDQELLAYA 427

RESULT 12  
US-10-166-984-75  
Sequence 75, Application US/10166984  
Publication No. US20040005562A9  
GENERAL INFORMATION:

APPLICANT: Croteau, Rodney et al.  
TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
FILE REFERENCE: 4630-62267  
CURRENT APPLICATION NUMBER: US/10/166,984  
CURRENT FILING DATE: 2002-06-10  
PRIOR APPLICATION NUMBER: US 09/866,570  
PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: US 09/457,046  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: US 09/411,145  
PRIOR FILING DATE: 1999-09-30  
NUMBER OF SEQ ID NOS: 76  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 75  
LENGTH: 435  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana

Query Match 18.9%; Score 448; DB 15; Length 435;  
Best Local Similarity 29.8%; Pred. No. 2e-38;  
Matches 137; Conservative 79; Mismatches 193; Indels 50; Gaps 15;

3 KIEVINSKHITKSTSTPIQPKYKLTLLDQTPPAYVPIVFFPIHDDFNLPQTLADL 62  
4 KLEVT--GKEVIKPA-SPSPDRQLQSLDLYCPIYVSTIFFYDLITSESEV--FSENL 58

63 ROALSETLTYPLSGRVKNNLYIDDFEEGVPYLEARVNCMDTDFLRKRLKIECLNEFVPI 122  
59 KLSLSETLSRFYPLAGRIE-GLSISCNDEGAVFTEARTDILLPDLFRLNLTDSLSGFLPT 117

123 KPFSMEALSDEYPLLGQVNVF--DSGIAIGVSVSHKLIIDGGTADCFKLGWAVFRGCR 180  
118 ---LAAGESPAWPLLSVKYTFFGSGGVAVSVSHKICDIAASLVTFVKWATTTAKGK 174

181 EN-----IHPSLSEALLPFRD---DLPEKYVDQMEALWFAKGVATRRFVGV 228  
175 SNSTIEFAATTIYPPPHMVEQFPSTDSNITSKYV-----LKRFPVFP 220

229 KAISIOEAKSESVPKSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAAQAVALRTR 288  
221 SKIAELKHKAASESVVPVTRVEAINSLIWR---CARNSRSNLLIPROAVMWAQMDIRLR 277

289 MNMTVLDNATGNFWAQAILELSHTTPEISDLKCDLVNLLSGVKOCNGDYFTFKG 348  
278 IPSSVAPKDVIGNL---QSGFSLKDAE--SEFEIPEIVATPRKNKRVNEMIKESLQ 331

349 KEGYGRMCYLDPORTMSMEPADYILFSSWTNF-FNPLDFGWTGTSWIGVAGKIESAS 407  
332 NT-IGQSLLSLMAETVSESTE--IDRYIMSSWCRKPFYEVDFGSGSPVWVGYSHTIYN 388

408 CKFIILVPTCCSGGIRAWNLBEKMWALEQDPHFLALA 446  
389 MVGVLLDSKGDGVAEWISLPEEDMSVFVDDQELLAYA 427

RESULT 13  
US-10-167-603C-7  
Sequence 7, Application US/10167603C  
Publication No. US20030226161A1  
GENERAL INFORMATION:

APPLICANT: ZENK, Meinhard H.  
APPLICANT: KITCHAN, Toni  
APPLICANT: GROTHE, Torsten  
TITLE OF INVENTION: Salutaridinol 7-O-Acetyltransferase and derivatives  
FILE REFERENCE: 67529  
CURRENT APPLICATION NUMBER: US/10/167,603C  
CURRENT FILING DATE: 2002-06-11  
PRIOR APPLICATION NUMBER: EP 0114122.3  
PRIOR FILING DATE: 2001-06-11  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 7  
LENGTH: 474  
TYPE: PRT  
ORGANISM: Papaver somniferum

Query Match 18.5%; Score 438; DB 12; Length 474;  
Best Local Similarity 27.9%; Pred. No. 2.6e-37;  
Matches 140; Conservative 87; Mismatches 170; Indels 104; Gaps 21;

1 MEKIEVSINKHITKST--SSTPIQPKYKLTLLDQTP-PAYVPIVFFPIHDDFNLPQTLADL 49  
4 MYSAAVEVISEKITKPTTPTPSQLKNFNLSLDQCFDYIYVPIILFPATAANSTGSSN 63

50 DHDNLPQTLADLROALSETLTYPLSGRVKNNLYIDDFEEGVPYLEARVNCMDTDFLR 109  
64 HHD-----DLDLKSSLSKTLVHFYPMAGRMIDNILDVCHDQGINFYKVKIRGKACEFMS 118

110 LRKIECLNEFVPIKPFPSMEALSDEYPLLGQVNVF--SGIAIGVSVSHKLIIDGGTADCF 168  
119 QPDVP-LQSLPSEVWS---ASVPKEALVIVQVNMEDCGGTACSSVSHKIAADATMTTF 174

159 LKSGWAVFRGCR-----ENIHPSLSEALLPFRD-----PEKY 205  
175 IRSMASITKTSRSGSGSTAATVDQKLIPSF-DSASLFPSPSERLTSFGSGSEIPEFSTPPT 233  
206 VDQMEALWFAKGVATRRFVGVKAISIQ-----DEAKSESVKPFVHVTGFLM 257

Db 234 EDD-----KTVSKRFVDFPAKITSVREKQVLMDHNYKR---RQTRVVFVTSLIW 281  
QY 258 KHLIAASRALTSSTTSLSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316  
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAFLPATTTT 333  
QY 317 -----PEISDLKCLDVLNLLNGSVKQCNGDYFETFKGEGYGRMCCEYLDFO 362  
Db 334 TNAVNTINSTSESQVVLHELHDFIAQMRSEIDKVKGDKGSLKVIQNFASGHD----- 388  
QY 363 RTWSSMEPADDI-----YLFSSWTFNFPNPLDFGWRGRTSWIGVAGKIE-SASCKFIILVPT 416  
Db 389 ---ASIKKINDVEINFWISSWCRMGLEYIDFGWKPFIWTVDPNKPKNKCF--MNDT 443  
QY 417 QCGSGIEAWNLLEBKMALE 437  
Db 444 KCGEGIEVWASFLDDMAKFE 464

RESULT 14  
US-10-167-603C-14  
; Sequence 14, Application US/10167603C  
; Publication No. US20030226161A1  
; GENERAL INFORMATION:  
; APPLICANT: ZENK, Meinhard H.  
; APPLICANT: GROTHE, Torsten  
; TITLE OF INVENTION: Salutaridinol 7-O-Acetyltransferase and derivatives  
; TITLE OF INVENTION: thereof  
; FILE REFERENCE: 67529  
; CURRENT APPLICATION NUMBER: US/10/167,603C  
; PRIOR FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: EP 01114122.3  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Papaver somniferum  
US-10-167-603C-14

Query Match 18.5%; Score 438; DB 12; Length 474;  
Best Local Similarity 27.9%; Pred. No. 2.6e-37;  
Matches 140; Conservative 87; Mismatches 170; Indels 104; Gaps 21;

QY 1 MEKIEVSINSKHTIKPST-SSTPLQPYKLTLLDQTP-PAYVPIVFFVPIIT----- 49  
Db 4 MYSAAVEVISKETIKPTTPSQKNFNLSDQCFPLYVYVPIILFPATAANSTGSSN 63  
QY 50 DHDENLPTQLADLQALSETLTYPLSGRVKNLYIDDFEFGVPLYEARVNCMDTDFLR 109  
Db 64 HED-----DLDLKSSLSKTLVHEYPWAGRMIDNILDCHDQGINFYKVKIRGKCEFMS 118  
QY 110 LRKIECLNEFVPIKPFSEALSDERYPLLGQVNVDF-SGIAIGVSVSHKLIIDGTDACF 168  
Db 119 QPDVP-LSQLLPSEWVS---ASVPEKALVIVQVNMDFCGGTACISSVSHKIAADAATMSTF 174  
QY 169 LKXSGAVRGR-----ENIHPSLSRAALLFPDRL-----PEKY 205  
Db 175 IRSWASTTKTSRSGSTAANTDQKLIPSF-DSASLFPSERLTSPSGMSEIPFSTPDET 233  
QY 206 VDOMEALWFAGKVKVATRRFVGVKALSIQ-----DEAKSEVPKPSRVHVAVTGLM 257  
Db 234 EDD-----KTVSKRFVDFPAKITSVREKQVLMDHNYKR---RQTRVVFVTSLIW 281  
QY 258 KHLIAASRALTSSTTSLSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316  
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAFLPATTTT 333  
QY 317 -----PEISDLKCLDVLNLLNGSVKQCNGDYFETFKGEGYGRMCCEYLDFO 362

Db 334 TNAVNTINSTSESQVVLHELHDFIAQMRSEIDKVKGDKGSLKVIQNFASGHD----- 388  
QY 363 RTWSSMEPADDI-----YLFSSWTFNFPNPLDFGWRGRTSWIGVAGKIE-SASCKFIILVPT 416  
Db 389 ---ASIKKINDVEINFWISSWCRMGLEYIDFGWKPFIWTVDPNKPKNKCF--MNDT 443  
QY 417 QCGSGIEAWNLLEBKMALE 437  
Db 444 KCGEGIEVWASFLDDMAKFE 464

RESULT 15  
US-10-167-603C-16  
; Sequence 16, Application US/10167603C  
; Publication No. US20030226161A1  
; GENERAL INFORMATION:  
; APPLICANT: ZENK, Meinhard H.  
; APPLICANT: KUTCHAN, Toni  
; APPLICANT: GROTHE, Torsten  
; TITLE OF INVENTION: Salutaridinol 7-O-Acetyltransferase and derivatives  
; TITLE OF INVENTION: thereof  
; FILE REFERENCE: 67529  
; CURRENT APPLICATION NUMBER: US/10/167,603C  
; PRIOR FILING DATE: 2002-06-11  
; PRIOR APPLICATION NUMBER: EP 01114122.3  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Papaver somniferum  
US-10-167-603C-16

Query Match 18.5%; Score 438; DB 12; Length 474;  
Best Local Similarity 27.9%; Pred. No. 2.6e-37;  
Matches 140; Conservative 87; Mismatches 170; Indels 104; Gaps 21;

QY 1 MEKIEVSINSKHTIKPST-SSTPLQPYKLTLLDQTP-PAYVPIVFFVPIIT----- 49  
Db 4 MYSAAVEVISKETIKPTTPSQKNFNLSDQCFPLYVYVPIILFPATAANSTGSSN 63  
QY 50 DHDENLPTQLADLQALSETLTYPLSGRVKNLYIDDFEFGVPLYEARVNCMDTDFLR 109  
Db 64 HED-----DLDLKSSLSKTLVHEYPWAGRMIDNILDCHDQGINFYKVKIRGKCEFMS 118  
QY 110 LRKIECLNEFVPIKPFSEALSDERYPLLGQVNVDF-SGIAIGVSVSHKLIIDGTDACF 168  
Db 119 QPDVP-LSQLLPSEWVS---ASVPEKALVIVQVNMDFCGGTACISSVSHKIAADAATMSTF 174  
QY 169 LKXSGAVRGR-----ENIHPSLSRAALLFPDRL-----PEKY 205  
Db 175 IRSWASTTKTSRSGSTAANTDQKLIPSF-DSASLFPSERLTSPSGMSEIPFSTPDET 233  
QY 206 VDOMEALWFAGKVKVATRRFVGVKALSIQ-----DEAKSEVPKPSRVHVAVTGLM 257  
Db 234 EDD-----KTVSKRFVDFPAKITSVREKQVLMDHNYKR---RQTRVVFVTSLIW 281  
QY 258 KHLIAASRALTSSTTSLSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316  
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAFLPATTTT 333  
QY 317 -----PEISDLKCLDVLNLLNGSVKQCNGDYFETFKGEGYGRMCCEYLDFO 362  
Db 334 TNAVNTINSTSESQVVLHELHDFIAQMRSEIDKVKGDKGSLKVIQNFASGHD----- 388  
QY 363 RTWSSMEPADDI-----YLFSSWTFNFPNPLDFGWRGRTSWIGVAGKIE-SASCKFIILVPT 416  
Db 389 ---ASIKKINDVEINFWISSWCRMGLEYIDFGWKPFIWTVDPNKPKNKCF--MNDT 443  
QY 417 QCGSGIEAWNLLEBKMALE 437  
Db 444 KCGEGIEVWASFLDDMAKFE 464

Tue Jul 6 14:13:41 2004

us-09-857-518a-6.rapb

Page 8

Search completed: July 6, 2004, 13:48:41  
Job time : 50.964 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:50 ; Search time 19.1722 Seconds  
(without alignments)  
2267.788 Million cell updates/sec

Title: US-09-857-518A-6

Perfect score: 2373

Sequence: 1 MEKIEVSINSKTIKPTSS.....MAMLEQDPHELALASPKTILI 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191536 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: Pir1:\*

2: Pir2:\*

3: Pir3:\*

4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	498	21.0	436	D86378	protein F21J9.8 [i
2	485	20.4	446	D71418	hypothetical prote
3	448	18.9	435	E71418	hypothetical prote
4	262	11.0	430	T46216	hypothetical prote
5	242	10.2	446	T10711	anthranilate N-ben
6	242	10.2	461	T00918	hypothetical prote
7	234.5	9.5	461	T51458	hypothetical prote
8	233.5	9.8	445	T10719	anthranilate N-ben
9	232	9.8	442	T10718	anthranilate N-ben
10	228.5	9.6	451	T00527	hypothetical prote
11	228	9.6	464	G86453	YOP812R.39. homol
12	225	9.5	445	T10717	anthranilate N-ben
13	222	9.4	572	T01056	hypothetical prote
14	219	9.2	452	G84823	probable anthocyan
15	217	9.1	454	T45611	N-hydroxycinnamoyl
16	215	9.1	451	T01140	hypothetical prote
17	209	8.8	456	T45610	proanthranilate N-
18	207	8.7	464	T45612	N-hydroxycinnamoyl
19	200.5	8.4	450	T45573	anthranilate N-hyd
20	196	8.3	443	T45574	anthranilate N-hyd
21	184	7.8	459	T45576	anthranilate N-hyd
22	180	7.6	450	T48479	hypothetical prote
23	178.5	7.5	461	H84644	hypothetical prote
24	177.5	7.5	458	T06313	hypothetical prote
25	177	7.5	421	T05583	CER2 protein - Ara
26	175	7.4	455	T09666	probable anthranil
27	173.5	7.3	460	T03274	hsr201 protein, hy
28	172	7.2	428	T48008	hypothetical prote
29	166.5	7.0	436	T00908	hypothetical prote

protein F21J9.8 [i  
hypothetical prote  
hypothetical prote  
protein F1K23.12 [i  
anthranilate N-ben  
anthranilate N-ben  
hypothetical prote  
hypothetical prote  
10-deacetylbaecat  
taxadienol acetyl  
hypersensitivity r  
hypothetical prote  
hypothetical prote  
alanine-tRNA ligas  
MAPK delta-1 prot  
trichothecene 3-O-  
bacitracin synthet

30 163 6.9 465 2 D86166  
31 159.5 6.7 648 2 T45575  
32 159 6.7 433 2 H84826  
33 158.5 6.7 480 2 H8411  
34 157 6.6 475 2 T45961  
35 154.5 6.5 450 2 T02368  
36 154 6.5 469 2 C86170  
37 148.5 6.3 440 2 T52320  
38 144 6.1 439 2 T52321  
39 142 6.0 447 2 T45653  
40 139 5.9 426 2 S68660  
41 132.5 5.6 414 2 T13436  
42 102.5 4.3 900 2 S73748  
43 101 4.3 886 2 T48544  
44 100 4.2 451 2 T43733  
45 99.5 4.2 2607 2 T31678

ALIGNMENTS

RESULT 1  
D86378 protein F21J9.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D86378  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alcorco  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.  
ansen, N.P.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016713; PMID:11130712  
A:Accession: D86378  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-436 <STO>  
A:Cross-references: GB:AB005172; NID:g9743331; PIDN:AAF97955.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F21J9.8  
A:Map position: 1

Query Match 21.0%; Score 498; DB 2; Length 436;  
Best Local Similarity 30.9%; Pred. No. 9.5e-34;  
Matches 145; Conservative 90; Mismatches 194; Indels 50; Gaps 19;

QY 1 MEKIEVSINSKTIKPTSSSTP--LQPYKTLIDQLTPPAYVPVFPYPTDHD---FNL 55  
Db 1 MEK-NVEILSREIVKVS-SPTDDKRIILSLDILSSPMYTGALLEAYADPQNLGFS 58  
QY 56 POTLADLRQALSETLIYPLSLGRVKNLYIDFEEGVPYLEARVNCMDTDFLRKIEC 115  
Db 59 EETSLLKLSKSLKTLPIFYPLAGRIIGS-FVECNDEGAVFEARVDHLLSEFLKCPVES 117  
QY 116 LNFVPIKPSMAISDERYPILGVQVNFED-SGIALGVSVSHKLDGGTADCFLKSWGA 174  
Db 118 LELLIPVEAKSREAVT---WPVLLICANFFSCGLVITICVSHKITDATSLAMFIRGAE 174  
QY 175 VFSGRENITHPSLSRAALLFPREDLPKYPVQMEALWPAKGKVAITRRVFGVKAISSI 234  
Db 175 SSRGLGITLI-PSFTASEVPFKPDLDELPSKPMDKKEV--BEMSCVTKRFVFDASKIKKL 231  
QY 235 QDEAKSEVPKPSRVHATVGLWKHLIAASRALTSCTTSLSLAAQAVNLRTMNMETV 294  
Db 232 RAKASRLNVKNPIRVEAVTALFWRCVTKVSRLL---SSLTPRTSVLQILMNLGRK--VDSL 286  
QY 295 LDNATGNLFWN-----AQAILLSHTTPEISDLKCLDLVNLNGSVKQCGNDYFETFK- 347

Db 287 CENTIGNMLSLMLKNEEAIERIQDVDBIRAK--EIFSLNCKEMSKSSSRIFELLE 344  
QY 348 -GREGYGRMCEYLDLDFORTMSSMEPADYILFSSWTFN-FNPLDFGWRGTSMIGVAGKIES 405  
Db 345 IGKV-YGRGNEM-----DLWMSNWKGLGYDADGMRKPVW--VTGRGTS 387  
QY 406 ASCKFIILVPTQCGSGIEAWNVNLEEKAMLEQDPHFLALAS--PKTLI 452  
Db 388 HFKNMLLIDTKDGGIEAMITLIEQMSIFECQELLESASLNPPVLI 436  
RESULT 2  
E71418  
hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 20-Jun-2000  
C:Accession: D71418  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113; PMID:9461215  
A:Accession: D71418  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-446 <BRV>  
A:Cross-references: GB:297338; NID:g2244870; PID:g2244896  
C:Genetics:  
A:Map position: 4COP9-4G3845  
C:Superfamily: Arabidopsis CER2 protein  
Query Match 20.4%; Score 485; DB 2; Length 446;  
Best Local Similarity 29.1%; Pred. No. 1.2e-32;  
Matches 143; Conservative 91; Mismatches 165; Indels 92; Gaps 23;  
QY 4 IEVSINSKHTIKSTSTP--LQPKLTLLDQLTERAVPIVFFYPTDHD-FNLPTLA 60  
Db 6 MKVETTSKEILIKS-SPTNNLQTLQISYDHIUPLPPVITVAFLEY--TKNDLISQKTS 62  
QY 61 DLQALSETLTYPLSGRVKNLYIDDFEGVYPLARY-NCMDTDFLRKTECLNEF 119  
Db 63 KLKTSLETITKFPYLAGRI-TGVTVCTDGBAIFVDARVNNCPTEFLKCPDFDALQOL 121  
QY 120 VPIKPFMEALSIDERYELLGWVWFV-D-SGIAIGVSVSHKIDGTDADCFKSGAVFRG 178  
Db 122 LPDWDVNDPYVAATWELLVAKATYFCGGNAGICITHKIADAASISTFIRSWAATARG 181  
QY 179 CREN---IIHPSLSEALLPPRRDLDPEKYVDQMEALWFAKGYA-TRRFVFGVKAIS 234  
Db 182 --ENDAAAMESPVFAGANFPANEAFLPADEQ-----AKRSITKRFVFEASKVEDL 234  
QY 235 QDEAKS-ESVPKPSRVHVTGFLWKHLIAASRALTSGTSTRLSIAAQVNLRTMMET 293  
Db 235 RTAAASSETVDQPRVESVTALINKCFVASSK-----TTTCDHKVLVQLANLRSKI-PSL 288  
QY 294 VLDNATGNLFWA-----QALLESHTTPTSDLKCD----- 326  
Db 289 LQESSIGNLAFSSVVLISIGRGEVKIEAVADRLRKKBELGTVLIDEGSGDSSSMIGSK 348  
QY 327 LVNLLNGSVKQCMGYETFTKGEYGRMCEYLDLDFORTMSSMEPADYILFSSWTFNFP 386  
Db 349 LANLM-----LTNYSRL-----SYTHPE-----YTSSCKL--P 377  
QY 387 L---DFNGRTSMIGVAGKIESASCKFIILVPTQCGSGIEAWNVNLEEKAMLEQDPHFL 443  
Db 378 LYEASFQWDSFVW--VGVNVSPVLGNLAWLIDSKDGGIEAFVTLPEENNMSFQNPPELL 435  
QY 444 ALA--SPKTLI 452

Db 436 AFATMNPVLV 446  
RESULT 3  
E71418  
hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
C:Accession: E71418  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113; PMID:9461215  
A:Accession: E71418  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-435 <BRV>  
A:Cross-references: GB:297338; NID:g2244870; PID:e326926; PID:g2244897  
C:Genetics:  
A:Map position: 4COP9-4G3845  
Query Match 18.9%; Score 448; DB 2; Length 435;  
Best Local Similarity 29.8%; Pred. No. 1.5e-23;  
Matches 137; Conservative 79; Mismatches 193; Indels 50; Gaps 15;  
QY 3 KIEVSINSKHTIKSTSTP--LQPKLTLLDQLTERAVPIVFFYPTDHD-FNLPTLA 62  
Db 4 KLEVT--GKEVIKPA-SFSPDRDLQSLIDLYCGIYVSTIFFYDLITESSEV--FSENL 58  
QY 63 QALSETLTYPLSGRVKNLYIDDFEGVYPLARYNCMDTDFLRKTECLNEFVPI 122  
Db 59 KLSLSETLSRFPYLAGRIE-GLSISCNDEGAVFTEARTDLLLLPDLRLMLNDSLSGFLPT 117  
QY 123 KPSMEALSIDERYELLGWVWF--DSGIAIGVSVSHKIDGTDADCFKSGAVFRGCR 180  
Db 118 ---LAAGESPAMPDLJSVKVTFPGSGGVAVSVSVSHKIDIASLVTFPKDWATTTAKG 174  
QY 181 EN-----IIHPSLSEALLPPRRD---DLPEKYVDQMEALWFAKGVATRRFVFGV 228  
Db 175 SNSTIEEAETIYPPPSHMYEQPSTSDSNITSYV-----LKRFPPEP 220  
QY 229 KAISSIDEAKSVKPSRVHVTGFLWKHLIAASRALTSGTSTRLSIAAQVNLRT 288  
Db 221 SKIAELKHKAASESVVPVTRVEATVSLIWR---CARNSSRSNLLIPRQAVWQAMCDRLR 277  
QY 289 WMETVLDNATGNLFWAQAILELSHTTPTSDLKCDLVNLLNGSVKQCMGYETFTFKG 348  
Db 278 IPSVVAQDVIGNL-----QSGFSLKKDAB--SEFEIETVATFRKNERNVEMKESLQG 331  
QY 349 KEGYGRMCEYLDLDFORTMSSMEPADYILFSSWTFN-FNPLDFGWRGTSMIGVAGKIESAS 407  
Db 332 NT-IGQSLLSLMAETVSESTE--IDRYIMSSWCRCPPFEVDVFGSGSPVWVGYSHTIYDN 388  
QY 408 CKFIILVPTQCGSGIEAWNVNLEEKAMLEQDPHFLALA 446  
Db 389 MVGVVLIDSKEGDGEAVMSILPEEDMSVFDQQLLAYA 427  
RESULT 4  
T46216  
hypothetical protein T8P19.230 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 17-Mar-2000  
C:Accession: T46216  
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; Sa  
submitted to the Protein Sequence Database, December 1999

A;Reference number: Z23008

A;Accession: T46216

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-430 <CHO>

A;Cross-references: EMBL:AL13315

A;Experimental source: cultivar Columbia; BAC clone T0P19

C;Genetics:

A;Map position: 3

A;Introns: 144/3

A;Note: T0P19.230

C;Superfamily: Arabidopsis CER2 protein

Query Match 11.0%; Score 262; DB 2; Length 430;  
Best Local Similarity 24.5%; Pred. No. 5.7e-14;  
Matches 115; Conservative 72; Mismatches 176; Indels 106; Gaps 23;

```

QY 16 PPSSTPLQYKLTLLDQTPPPAYPIVFPY-----ITDHPNLQTLADLQALSTLT 71
DB 18 PPVSETPNGHYLSNLDQ-NIAIVKTLVYKGSRTNQESYV-----IKKSLSEVLV 70
QY 72 LYPLSGRV-----KNMLYIDDFEGVPLEARVNC---DMTDLRLKIECIENEFYPIKP 124
DB 71 HYPVAGRLTISPEGKIANVCTGEGVVVVEAEANCIGDITIKKALSENRMETLEKLYDVP 130
QY 125 FSEALISDERYPPLGGVQVNVFD-SGIAIGVSVSHKLDGGTADCFKLSWGAVERGCRENI 183
DB 131 GARNIL---EIPFVVQVYTFKGGFVGLGMSINMFEDGVAAEFLNSWCEMAKGLPLSV 187
QY 184 IHPSLSEAALL---FPPRDLPEKYVDQMALWPAK-----KVATRRFVFGVKAISIQ 235
DB 188 --PPFLDRTILGRNPKIEFPNHEFOIEDISTGKIYDEEKLYKSFLEPEKLEK 245
QY 236 DEAKSE-SVPKPSRVHATGFLKGLIAASGALTSGTSTRLSIAAQAQVNLRTM----- 289
DB 246 IMAIEENNNKVSFTQALTGFLKSRCEALR--FKPDQRYKLLFAADG---RSRFTPLRP 300
QY 290 -----NMETVLNDNATGNLFWAAQAILLSHTTPEISDLKCDLVNLLNGSVKQC 338
DB 301 QGCGNGIVLTGLVSSGELVGN-----PLSHSVGLVK-----RLVELVTDGFWRS 346
QY 339 NGDYFTFKGEGYGRMCEYLDQRTMSMEPAPDIYLFSSWTNF-FNPILDQFGWRTSW- 396
DB 347 AMDYFEV-----NRTSPSNAT---LLITSWSKLTLLKLDQFGNGEPVFS 387
QY 397 --IGVAGKTESACKPILLYPTCGSG-----TEAWNLSEKXAMLEQ 438
DB 388 GPVGLPGR-----EVLFLP-----SGDDMKISINVFLGPTSAMVFEER 426

```

#### RESULT 5

T10711

anthranilate N-benzoyltransferase (EC 2.3.1.144) - clove pink

N;Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase

C;Species: Dianthus caryophyllus (clove pink)

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000

C;Accession: T10711

R;Yang, Q.; Watern, U.; Grimmig, B.

Plant Mol. Biol. 38, 1201-1214, 1998

A;Title: Anthranilate N-hydroxycinnamoyl/benzoyltransferase gene from carnation: Rapid e

A;Reference number: Z17092; MUID:99084770; PMID:9869425

A;Accession: T10711

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-446 <YAN>

A;Cross-references: EMBL:Z98758; NID:93288179; PIDN:CAB11466.1; PID:93288180

C;Function:

A;Description: catalyzes the synthesis of anthranilate

A;Pathway: phytoalexin biosynthesis

C;Keywords: acyltransferase; coenzyme A

Query Match 10.2%; Score 242; DB 2; Length 446;

Best Local Similarity 23.3%; Pred. No. 2.9e-12;

Matches 114; Conservative 88; Mismatches 185; Indels 102; Gaps 26;

```

QY 4 IEVSINSKHTIKSTSTPLQYKLTLLDQ--TPPAYPIVFPYPTD----- 51
DB 1 MSIQIKOSTWREFA-BETPNKSLWLSKIDMLIRFYSHTGAVLLYKQPDNNEDNIHSSS 59
QY 52 ---DFNLQTLADLQALSETLTLYPLSGRVKNN---LYIDDFEGVPLEARVNCIDMT 105
DB 60 MYFDANI-----LIEALSALVPIYYPMAGRLKINGDRYELDCNAEGALFVEAESHVLE 113
QY 106 DFLRLAKIECINE-FVPIKPFSEAIISDERYPPLGGVQVNVFD-SGIAIGVSVSHKLDGG 163
DB 114 DFGDFRPNDLHRVMVPTCDYS-KGISS--FPLIMVQLTRFCGGSIGFQAQHEHACDGM 170
QY 164 TADCFKLSWGAVERGCRENI--IHPSLSEAALLFEPF-----DDLPEKYVDQM 209
DB 171 SHEEFNNMWARIKAGLLPALPEVHRYLHLRLRNPQIKVTHSQEPFVPSLPNELLD-- 228
QY 210 EALWFAGKVKVATRR-FVFGVKAISIQD--BAKSEVFPKPSRVHATGFLMKHLIAASRA 266
DB 229 -----GKTNKSQTLFKLSREQINTLKKLDLSSNTTTLSTYEVVAGHVWRS-VSKARG 281
QY 267 LTSGTSTRLSIAAQAQVNLRTMNMETVLNDNATGNLFWAAQAILLSHTTPEISDLKCD 326
DB 282 L-SDHEEIKL---IMPVDGRSRINNPSPKGYCGN-----VFLAVCTATVGDLS-CN 329
QY 327 LVNLLNGSVKQCGDYPTETFKGEGYGRMCEYLDQRTMSMEPAPDI----- 374
DB 330 PLTDTAGKVQ-----EALKGLDD-----DYL--RSAIDHTESKPLPFPYMSPEKTL 375
QY 375 ---YLFSSWTNF-FNPILDQFGWRTSWIGVAGKTESACKPILLYPTCGSG-IEAWNL 429
DB 376 YPNVLNWSGRIPYQAMDQFGSGPTFGISNIFYDGC---FLIPSQNGDSMTLAINLF 432
QY 430 EEXAMLEQ 433
DB 433 SSHLSLFFK 441

```

#### RESULT 6

T00918

hypothetical protein F21B7.32 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse ear cress)

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999

C;Accession: T00918

R;Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Co

eologis, A.; Ecker, J.R.

submitted to the EMBL Data Library, January 1998

A;Description: Genomic sequence for Arabidopsis thaliana BAC F21B7.

A;Reference number: Z14208

A;Accession: T00918

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-461 <SHI>

A;Cross-references: EMBL:AC002560; NID:92618677; PID:92809263; GSPDB:GN00059; ATSP:F21B

C;Genetics:

A;Gene: ATSP:F21B7.32

A;Map position: 1

Query Match 10.2%; Score 242; DB 2; Length 461;

Best Local Similarity 25.4%; Pred. No. 3e-12;

Matches 120; Conservative 76; Mismatches 169; Indels 108; Gaps 27;

```

QY 4 IEVSINSKHTIKSTSTPLQ-----YKLTLDQLTPPAYPIVFPYPTDHDNLPQ 57
DB 15 IPTVINOQFIVHFS-SPTPANQSPHSLYSLNLDODIIGARVFTPSVFPYSTNRESF-- 71
QY 58 TLADLQALSETLTLYPLSGRVK--NNLYIDDD---EEGVPLEARVNCIDMTQFLRLK 112
DB 72 VLKELQALSELVLPYPLSGRLREVNGKLEVFEGEQGLMVNSANSSMDLAD----- 125
QY 113 IECLNEFVPIKPFSEAI--SDERY-----PILGVQVNVFD-SGIAIGVSVSHKLDGG 162

```

Db 126 ---LGLDTPVNPAMPLIFRNPGSEAYKILEMPLLIAQVTFCTCGFSLGRLCHCICDG 182  
Qy 163 GTADCLKSGAVRGCRGNIHPS-----LSEALLFPDRDLP-EKYVDQWEA 211  
Db 183 FGAMQFSGWAATAK-TGKLIADPEPVNDRTFKRPPPMVKYPHHEYLPIERSNLTN 241  
Qy 212 LWFAGKKV-----ATRRFVFGVKAISIQDEAKSESVPKPSRV-----HAVTGFLEKHLIA 262  
Db 242 LMDT-KPLQCYRISKFCQVKSIAQED-----PTLVCSSTDAMAHAHWRSWV- 290  
Qy 263 ASRALTSGTSTRLSIAQAQVNLTRMNETVLDNATGNLFWAAQAIL-----ELSH 314  
Db 291 --KALDVKPLDYNLRLETF-SYVNRLETLKLRKGFYGVVVCACAMSSVESLINDLSLK 347  
Qy 315 TTPRISDLKCLDLVNLNGSKQCGNGDYFETFKGKGVRMCEYLDQRTWSSMEPAPDI 374  
Db 348 TTRLVQDARL-----RVSEYILRS-----MVDIVDWKRP-KRLEFGKGL 385  
Qy 375 YLFSSWTNP--FNPLDFGWRGTSWIGVAGKIE-SASCKFIILVPTQCGSGIEA 424  
Db 386 TI-TOWTRFEMETADFGWCKPVY--AGPIDLRPTQVCVLLPQ---GGVES 431

RESULT 7  
T51458  
hypothetical protein K10A8\_20 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C;Accession: T51458  
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew  
submitted to the Protein Sequence Database, August 2000  
A;Reference number: Z25394  
A;Accession: T51458  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-461 <SAT>  
A;Cross-references: EMBL:AL391151  
A;Experimental source: cultivar Columbia; BAC clone K10A8  
C;Genetics:  
A;Map position: 5  
A;Introns: 147/3  
A;Note: K10A8\_20

Query Match 9.9%; Score 234.5; DB 2; Length 461;  
Best Local Similarity 24.7%; Pred. No. 1.3e-11;  
Matches 106; Conservative 69; Mismatches 178; Indels 77; Gaps 20;

Qy 39 YVPLVPEY---PITDHPNLPTLIADIRQALSETLTYLPLSGRVK-----NNLYIDDFEE 91  
Db 42 HIPTIFPYRNPNTNSD-----PVAVIRRALAETLVYYPFAGRLREGPNRKLAVDCTGE 96  
Qy 92 GVPYLEARVNCMDTDFLRKIE---CLNEFVPIKPFMEASISDE-RYPLIGVQVNVFD 146  
Db 97 GVLFIEADADVLVEFEKDALKPPFCPELL-----FNVEGCEMLNTPMLMQVTRLK 152  
Qy 147 -SGIAIGVSVSHKLIIDGTDADCLKSGAVRGCRGNIHPS-----SLSEALLFPDRD 200  
Db 153 CGGFIFAVRINHAMS DAGGLTLFLKTCFVRGVHYAFTVAPVWERHLLSARVLL---RVT 209  
Qy 201 LPEKYVDQWEALFAGKVVATR-----FVGVKAISIQDEAKSESVPKPSRVNAV 253  
Db 210 HAHEYDEMPAI---GTEIGSRDNLVGRSLFPQFCMSAIRLLPPLNLSVNSNTNMEMLT 266  
Qy 254 GFLWKHLIAASRALTSGTSTRLSIAQAQVNLTRMNETVLDNATGNLFWAAQAILLS 313  
Db 267 SPLWRYETIALR--PDQDKENRLLI---IVNARSKLNPPPLPRGYGNAPFVVAIA--- 318  
Qy 314 HTTPEISDLKCLDLVNLNGSKQCGNGDYFET-----FKGKGVRMCEYLDQRTWSS 367  
Db 319 -TANELTKKPLESALRLIKAKSSVTEYMRSLADIAMVIGRPSFSIDGAYL----- 369  
Qy 368 MEPAPDIYLFSSWTFNFPNPLDFG-WGRTSW--IGVAKIESASCKFIILVPTQCGS-GIE 423

Db 370 ---VSDVRIFAD-----IDFGIWGKPVYGGIGTAGVEDLFGASFYVSEKRNGEIGIV 419  
Qy 424 AAWNLEEEKM 433  
Db 420 VPVCLPEKAM 429

RESULT 8  
T10719  
anthranilate N-benzoyltransferase (EC 2.3.1.144) [clone pchcbr3] - clove pink  
N;Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase  
C;Species: Dianthus caryophyllus (clove pink)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C;Accession: T10719  
R;Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.  
Plant Mol. Biol. 35, 777-789, 1997  
A;Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA: a  
lus L.  
A;Reference number: Z17095; MUID:98088004; PMID:9426598  
A;Accession: T10719  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-445 <YAN>  
A;Cross-references: EMBL:Z84571; NID:g2239090; PIDN:CA806538.1; PID:g2239091  
C;Function:  
A;Description: catalyzes the synthesis of anthranilate  
A;Pathway: phytoalexin biosynthesis  
C;Keywords: acyltransferase; coenzyme A

Query Match 9.8%; Score 233.5; DB 2; Length 445;  
Best Local Similarity 23.5%; Pred. No. 1.5e-11;  
Matches 114; Conservative 92; Mismatches 189; Indels 91; Gaps 27;

Qy 4 IEVINSKHTIKSTSTPLQPYKLTLLDQL--TPPAVVPVVFYPIIDHDFNLPCITLAD 61  
Db 1 MSHIKOSTWVRPA-EETPNKSLMLSKIDMLIRTPSYHTGAVLTKOPDNNEDNIQSSS 59  
Qy 62 -----LQALSETLTYLPLSGRVKNN---LYIDDPEGVPVLEARNVNCMDTDFLR 111  
Db 60 MYFDANILIEALSALVPYYPFMAGRLKINDRYEIDCNGEGALFVEAESSHVLEDFGDFR 119  
Qy 112 KIECLNE-FVPIKPFMEASISDERYPILIGVQVNVFD-SGIAIGVSVSHKLIIDGTDADCL 169  
Db 120 PNDELHRVMVPTCDYS-KGISS--FPLIMVQLTRFCGVSIGFAQHHVCDRMSHFEFN 176  
Qy 170 KSMGAVFRG-----CFENLIH-----PSLSEALLFPD-RDDLPEKYVDQWEALWFA 215  
Db 177 NSWARIAGLLPALEFPVHDRLYLCLPNPQIKYTHSQFEPFVPSLPELDD----- 228  
Qy 216 GKQVATRR-FVFGVKAISIQDEAK-SESVKPSRVHVTGFLWKHLIAASRALTSGTTS 273  
Db 229 GKTSKQTLFKLSSEQINTLKQKLDWNTTTLSTYEVVAGHVWRS-VSKARGL-SDHEE 286  
Qy 274 TRLSIAQAQVNLTRMNETVLDNATGNLFWAAQAILSETTEISDLKCLDLVNLNG 333  
Db 287 IKL---IMPVDEGSRINNPSPKPGYCGN-----VVFLAVCTATVGDLE-ACNPLTDTAG 335  
Qy 334 SVKQCGNDYFETFKGKGVRMCEYLDQRTWSSMEPAPDI-----YLF 378  
Db 336 KVQ-----EALKGLDD-----DYL--RSAIDHTSKPDLIPVPMGSPEKTLIYNVLVN 381  
Qy 379 SWTNF-FNPLDFGWRGTSWIGVAKIESASCKFIILVPTQCGSG-IEAWNLEEEKMAM 436  
Db 382 SWGRIPYQAMDFGNGNPTFFGISNIFYDGC---FLIPSGNGSGMWTALNLFSSLSLF 438  
Qy 437 EQDPHF 442  
Db 439 KK--HF 442

RESULT 9

T10718  
anthranilate N-benzoyltransferase (EC 2.3.1.144) [clone pchcbr3] - clove pink (fragment

N;Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase  
C;Species: *Dianthus caryophyllus* (clove pink)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C;Accession: T10718  
R;Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.  
Plant Mol. Biol. 35, 777-789, 1997  
A;Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate L.  
A;Reference number: Z17095; MUID:98088004; PMID:9426598  
A;Accession: T10718  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-442 <YAN>  
A;Cross-references: EMBL:Z84384; NID:G2239084; PIDN:CAB06428.1; PID:G2239085  
C;Function:  
A;Description: catalyzes the synthesis of anthranilate  
A;Pathway: phytoalexin biosynthesis  
C;Keywords: acyltransferase; coenzyme A

Query Match 9.8%; Score 232; DB 2; Length 442;  
Best Local Similarity 23.6%; Pred. No. 1.9e-11;  
Matches 109; Conservative 82; Mismatches 173; Indels 98; Gaps 25;

QY 14 IKPST-----SSTPLQPKYLTLDQL--TPAYVPVIVFFYPITDH-----DF 53  
DB 2 IKQSTWVPAEETPKSLWLSKIDMILRTPYSHGCAVLTKQPDNEDNHPSSMYFDA 61  
QY 54 NLPQTLDLQALSETLTYPLSGRVKN---LYIDDFEGVPYLEARVNCMDTDFLR 110  
DB 62 NI-----LIEALSALVPVPPMAGRLKINGDRYEIDCNAGALFVEAESHVLEDFGDF 115  
QY 111 RKTCLNE-RVPKPPSMEASDERYDILGVQNVFDP-SGIAIGVSVSHKLDGTADCP 168  
DB 116 RPNDLHRVMVPTCDYS-KGISS--FPLLMVQLTRFCGGVSGIQAQHHVDCGVAHFF 172  
QY 169 LKSGAVFRG-----CRNLIH-----PSLSEALLRPP-RDDLPEKYVDQMEALWF 214  
DB 173 NNSWARIAGLLPALEPHRYHLRPNPPOIKYSHSQEPFVPSLPNELLD----- 225  
QY 215 AGKQVATR-FVFGVKAISIQDEAK-SESVKPSRVHVTGFLMKHLIAASRALTSQTT 272  
DB 226 -GKTNKSQTLFLLSREQINTLKQKLDLSNNTRELSTVEVAGVWRS-VSKARGL-SDHE 282  
QY 273 STLSIAAQAVALTRMMETVLNATGNLFWAQAILEL-SHTTPEISDLKCDLVNLSGVK 329  
DB 283 EIKL---IMPVDGSRINNPSPKGCN-----VFLAVCTATVGDLSCNELTDAG 332  
QY 330 LLGSGVQKCGDYPETPKGEGYGRMCEYLDQFTMSM-----EPADIIY---LFSSW 380  
DB 333 KVQELAKGLDDYLRS-----AIDHTSKPGLPVPMGSPKTLVPPNVLNSW 380  
QY 381 TNF-FNPLDRGWRGTSVIGVAGKTESASCKFIILVPTQCGSG 421  
DB 381 GRIPYQAMDFGWSGPTFFGISNIFDQGC---FLIPSRDGDG 419

RESULT 10  
T00527  
hypotheical protein At2g19070 [imported] - Arabidopsis thaliana  
N;Alternate names: hypotheical protein T20K24.8  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 02-Feb-2001  
C;Accession: T00527; B84572  
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, July 1997  
A;Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.  
A;Reference number: Z14167  
A;Accession: T00527  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-451 <ROU>  
A;Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176709  
A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84572  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-451 <STO>  
A;Cross-references: GB:AE002093; NID:g3176709; PIDN:AAD1025.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g19070; T20K24.8  
A;Map position: 2  
A;Introns: 322/1

Query Match 9.6%; Score 228.5; DB 2; Length 451;  
Best Local Similarity 22.6%; Pred. No. 3.9e-11;  
Matches 99; Conservative 70; Mismatches 182; Indels 87; Gaps 19;

QY 13 TIKPSTSTPLQ-----YKLTLDQLTPPAYWIVFFYPITDHFNLPQTLADLRQALS 67  
DB 5 TERKSYTIVFAEPTWSGRFPPLAEWCQWGTTHITPTLYFYDKPSSSQ-GNVVEILKTSLS 63  
QY 68 ETLLTYPLSGRV---KNLYIDDFEGVPYLEARVNCMDTDFLRKIECLNEFVPIK 123  
DB 64 RVLVHFTPMAGRLKWLPRGRFELNCNAGVZFIBAESEGLSDP-----KDFSTP 114  
QY 124 PFS--MEAIS---DERYPLLGQVNVFDP-SGIAIGVSVSHKLDGTADCPFLKSWGAVF 176  
DB 115 EFENLMPQVYKNFIETIPLFLAQVTKFGKGGISLVSVVSHAVDQGSALHLISEWGLA 174  
QY 177 RGRENIIHPSLSAAL-----LPPR-----DDLPEKYVDQMEALFAGKVKVATR 222  
DB 175 RGEPLTV-PFLDRKILWAGEPLFPFVSPPKFDEKFDQPPFLIGETDNVERKKKTTIV 233  
QY 223 RFVGVKAISIQDEAKSESVPK2---SRVHVTGFLMKHLIAASRALTSQTTSTLSI 278  
DB 234 MLPJSTSQLKRSKANGSKHSDSAKGFTRVETVGVHVR---CACKARGHSPQPTALGI 291  
QY 279 AAQAVNLRTRMMETVLNATGNLFWAQAILEL--SHTTPEISDLKCDLVNLSGVK 336  
DB 292 C-----IDTRSEMPPLRG-----YFGNATLDVVAASTSGELISNELGFAASLIKAIK 341  
QY 337 QCNGDYETFKGEGYGRMCEYLDQFTMSMEAPDIYLFSSWNTFF--NP----- 386  
DB 342 NVTINVTYMI-----GIYLNQKDLKFKQ---DLHALGSTGEPFYGNPNLGWVSWL 389  
QY 387 -----LDFGWGRTSWIG 398  
DB 390 TLENYGLDFGWSGKEFYTG 407

RESULT 11  
G86453  
YUP8H12R.39, homolog P9111.9 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: G86453  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K. ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G86453  
A;Status: preliminary  
A;Molecule type: DNA

[illegible]



Search completed: July 6, 2004, 13:39:30  
Job time : 21.1722 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:59 ; Search time 11.6195 Seconds  
(without alignments)  
2025.529 Million cell updates/sec

Title: US-09-857-518A-6

Perfect score: 2373

Sequence: 1 MEKIEVSINSKHTKESTSS.....MAMLEQDPHFLALSPKTLI 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	177.5	7.5	441	1	DBNT TAXCA
2	156	6.6	439	1	TSAT TAXCH
3	148.5	6.3	440	1	DBAT TAXCU
4	144	6.1	439	1	TSAT TAXCU
5	143	6.0	469	1	ANTA_GENTR
6	142.5	6.0	440	1	DBBT TAXCU
7	102.5	4.3	900	1	SYA_MYCPN
8	99.5	4.2	831	1	PLRL MEIGA
9	99.5	4.2	2607	1	PACB_BACLI
10	99	4.2	574	1	IPA4_SHIEL
11	99	4.2	1281	1	MDR3_CRIGR
12	96.5	4.1	855	1	XAB2_HUMAN
13	96.5	4.1	855	1	XAB2_MOUSE
14	96.5	4.1	855	1	XAB2_RAT
15	95.5	4.0	2227	1	POLG_HPAVH
16	95.5	4.0	2227	1	POLG_HPAVH
17	95	4.0	505	1	V4XG_RHLSN
18	95	4.0	928	1	NRPI_XENLA
19	94.5	4.0	2227	1	POLG_HPAVL
20	93	3.9	6486	1	TYCC_BREPA
21	92.5	3.9	741	1	LOXB_PHAVU
22	92.5	3.9	2226	1	POLG_HPAV2
23	92.5	3.9	2226	1	POLG_HPAV4
24	92.5	3.9	2226	1	POLG_HPAV8
25	91.5	3.9	397	1	V506_ROTGP
26	91.5	3.9	693	1	EX70_DROME
27	91	3.8	2230	1	POLG_HPAV5
28	90.5	3.8	1056	1	POL_BIV27
29	90	3.8	886	1	LEUR_YEAST
30	90	3.8	1276	1	MDR2_MOUSE
31	90	3.8	1278	1	MDR2_RAT
32	88	3.7	709	1	KSP1_SCHPO
33	87.5	3.7	397	1	V506_ROTGP

#### RESULT 1

DBNT TAXCA STANDARD; PRT; 441 AA.  
AC Q8LL69;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase (EC 2.3.1.-)  
GN DBTWTB).  
OS Taxus canadensis (Canadian yew).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Coniferopsida; Coniferales; Taxaceae; Taxus.  
ON NCBI\_TaxID=88032;  
RX MEDLINE=22103659; PubMed=12089320;  
RA Walker K., Long R., Croteau R.B.;  
RT "The final acylation step in taxol biosynthesis: cloning of the taxoid  
RT C13-side-chain N-benzoyltransferase from Taxus.";  
PL Proc. Natl. Acad. Sci. U.S.A. 99:9165-9171(2002).  
CC -!- FUNCTION: Catalyzes the stereoselective coupling of the surrogate  
CC substrate N-debenzoyl-(3'RS)-2'-deoxytaxol with benzoyl-CoA to  
CC form predominantly one 3'-epimer of 2'-deoxytaxol.  
CC -!- PATHWAY: Taxol biosynthesis; last step.  
CC -!- SIMILARITY: Belongs to the plant acetyltransferase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC EMBL: AF466397; AM75818.1;  
DR InterPro; IPR003480; Transferase.  
DR Pfam; PF02458; Transferase; 1.  
KW Taxol biosynthesis; Transferase; Acyltransferase.  
SQ SEQUENCE 441 AA; 49041 MW; 1F0C20CFDA6D4A0C CRC64;

Query Match 7.5%; Score 177.5; DB 1; Length 441;

Best Local Similarity 24.3%; Pred. No. 2.3e-07;

Matches 108; Conservative 71; Mismatches 167; Indels 99; Gaps 24;

Qy 14 IKPSTSTPQPKYKLLDQLTPPAYVPIVFYPTDHDNLPTL-AD----LRQALSE 68  
Db 19 VAPSLPS-PRATVQLSVVDSLT----ICRGIFNTLL--VFNAPDNISADPKIIRALSK 71  
Qy 69 TLTLIYPLSGRVKN----NLVIDDFEGVPVLEARNCDMTDFLRRLKIECLN-----BFV 120  
Db 72 VLIVYFFPLAGRLSRKSGEIVECTGCGALFVEMAVEDTIS----VLRLDLDLNFPSFOQLV 128  
Qy 121 PIKFSNEATSDERYPLGCVQVNVFD--SGIAGVSVSHKLIDGGTADCFKSWGAVFRG- 178  
Db 129 FWHPLD-TAIED--LHLIVIQVTRFTCGGIAGVGTLPVSHVCDGRGAAQFVTALAEARGE 185



Matches 99; Conservative 62; Mismatches 158; Indels 101; Gaps 20;

QY 14 IKPSTSTPLQPKYKLTLLDQTPPAYVPIV-----PYPITDHDNPLPQTLLADLRQALSET 69

DB 17 VAPQPS-PKAFIQLSTLNL---PGVRENIFNLLVYNASDRVSDPAKV--IRQALSKV 71

QY 70 LTYPLSLGRVK-----NNLYIDFEGVGYLVARVNCMDTDLRLKIECLNEFVPIKPF 125

DB 72 LVYSPFAGRLRKENGDEVECTGEGALFVEAMDT-----LSVLGLDDYSP----- 121

QY 126 SWEAI-----SBERPLLGQVNVFD-SGIALGVSHKLIIDGTCADCFKLSGWAV 175

DB 122 SLEQLLFCPLPDDTIDSIHELIV-VQVTRFCGFGVGVSEFCHGICDGLGAGQFLIANGEM 180

QY 176 FRGRENIIHPSLSEAL---LFPDRDLP--EKYVDQMEALWFPAGKKVATRRFVGVKA 230

DB 181 ARG---EIKPS-SEPIWKELIKPEDPLRYEQYHFQILCPPSTFGKIVQGSIVITSET 235

QY 231 ISSIQEAKSESVPKPSRVHVTGFLWKHLIAASRAL-TSGTTSTRLSIAAQVNLRTM 289

DB 236 INCIKQCLRESKEFCFAFEVWSALAW---IARTALQIPHSENVKLIFF---AMDWKLF 289

QY 290 N-----METVLNATGNLFWMAQAIIELSHSTTPETISDLKLCOLVNLN 332

DB 290 NPPLSGYGNFVGTVCAMDNVDLSSGL---RVRIIKAKVSLNEHFTSTIVTPRS 346

QY 333 GSVKQCNQDYFETFKGKEGYRMCVYLDFORTMSMEPAPDIYLFSSWTNFFNPLDFGNG 392

DB 347 GSDSINYNIVGFDRLRG-----FDEVDPGNG 376

RESULT 4

TSAT TAXCU STANDARD; PRT: 439 AA.

AC Q9M6F0;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Taxadien-5-alpha-ol O-acetyltransferase (EC 2.3.1.162) (Taxa-

DE 4(20),11(12)-dien-5alpha-ol-O-acetyltransferase) (Taxadienol

DE acetyltransferase).

GN TAT.

OS Taxus cuspidata (Japanese view).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferopsida; Coniferales; Taxaceae; Taxus.

OX NCBI\_TaxID=99806;

RN [1]

RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.

RX MEDLINE=20132877; PubMed=10666320;

RA Walker K., Schoendorf A., Croteau R.B.;

RT "Molecular cloning of a taxa-4(20),11(12)-dien-5alpha-ol-O-acetyl

RT transferase cDNA from Taxus and functional expression in Escherichia

RT coli.";

RL Arch. Biochem. Biophys. 374:371-380(2000).

CC -I- CATALYTIC ACTIVITY: Acetyl-CoA + taxa-4(20),11-dien-5-alpha-ol =

CC CoA + taxa-4(20),11-dien-5-alpha-yl acetate.

CC -I- PATHWAY: Taxol biosynthesis; third step.

CC -I- SIMILARITY: Belongs to the plant acetyltransferase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DB EMBL; AF190130; AAF34254.1; --

DR PIR; T52321; T52321.

DR InterPro; IPR003480; Transferase.

DR Pfam; PF02458; Transferase; 1.

KW Taxol biosynthesis; Transferase; Acyltransferase.

SEQUENCE 439 AA; 49079 MW; 7095CAF119B8B8DFF CRC64;

Query Match 6.1%; Score 144; DB 1; Length 439;

Best Local Similarity 21.0%; Pred. No. 0.00016;

Matches 102; Conservative 78; Mismatches 188; Indels 118; Gaps 22;

QY 1 MEKIEVSIN-SKHTIKPSTSTPLQPKYKLTLLDQTPPAYVPIVFFYPITDHDNPLPOT 58

DB 1 MEKTDHVNLIKVMVQPS-PPLPKTTQLQSSIDLNL---PGVRGSIFNALLIYNASPSPTM 57

QY 59 LA-----DIQALSETLTYPLSLGRVK-----NNLYIDFEGVGYLVARVNCMDTDLR 109

DB 58 ISADPAKPIREALAKILVYPPFAGRLRETEGDEVECTGEGAMFLEA-----MAD--- 109

QY 110 LRKIECLNEFVPIK-----FSMEALSD-ERYPLLGQVNVFD-SGIALGVSHKLIID 161

DB 110 -NELSLVGDEDDNPFQQLLSLDLDTNFKDLSLVVQVTRFCGFGVGVSEFHHGVCD 168

QY 162 GGTADCFKLSGWAVFRGRENIIHPSLSEALFPDRDLPPEKYVDQMEALWF----- 214

DB 169 GRGAQFLKGLAEWARG-----EVKLSLEPIWNRRELKLDLPKYLQPFHFEFLR 217

QY 215 ---AGKKVATRRFVGVKALSSIQEAKSESVPKPSRVHVTGFLWKHLIAASRALSGT 271

DB 218 APSIVEKIVQTYFTIIDPETINIKQSVMECKEFCSSFEVASATW---IARTAFQIP- 273

QY 272 TSTRLSIAAQVNLRTMNMETVLD-----NATGNLFWMAQAIIELSHSTTPETISDLKLC 326

DB 274 -----ESEYKILFGMDNRNSFNPLPSGYGNSIGTACAV-----DN 311

QY 327 LVNLINGS-----VQKMGDYFETFKGKEGYRMCVYLDFORTMSMEPA----- 371

DB 312 VQDLLSGSLRLATMIKKSKVSLNDNPKSR-----AVKPSELDNNVNH 355

QY 372 PDYLFSSWTNF-ENPLDFGWRFTSWGAGKTESACK--FIILVPTQ-CGSIEAWVN 427

DB 356 ENVVAFADMSRLGDEVDVFGWNAVSVSPVQOSALAMQNYFLFXPSKPKDGIKILMF 415

QY 428 LEEKM 433

DB 416 LPLSKM 421

RESULT 5

ANTA\_GENTR STANDARD; PRT: 469 AA.

AC Q9ZWR8;

DT 15-MAR-2004 (Rel. 43, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Anthocyanin 5-aromatic acyltransferase (EC 2.3.1.153) (SAT).

OS Gentiana triflora (Clustered gentian).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC lamids; Gentianales; Gentianaceae; Gentiana.

OX NCBI\_TaxID=55190;

CC [1]

CC SEQUENCE FROM N.A., SEQUENCE OF 68-89; 249-262; 395-404 AND 442-459,

CC CHARACTERIZATION, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND

CC SUBCELLULAR LOCATION.

CC TISSUE=Petal;

CC MEDLINE=9907837; PubMed=3881162;

CC Fujiwara H., Tanaka Y., Yonekura-Sakakibara K., Fukuchi-Mizutani M.,

CC Nakao M., Fukui Y., Yanaguchi M., Ashikari T., Kusumi T.;

CC "cDNA cloning, gene expression and subcellular localization of

CC anthocyanin 5-aromatic acyltransferase from Gentiana triflora.";

CC Plant J. 16:421-431(1998).

CC -I- FUNCTION: Transfers hydroxycinnamic moieties to the glucosyl

CC groups of anthocyanin.

CC -I- CATALYTIC ACTIVITY: Hydroxycinnamoyl-CoA + anthocyanidin-3,5-

CC diglucoside = CoA + anthocyanidin 3-glucoside-5-

CC hydroxycinnamoylglucoside.

CC -I- ENZYME REGULATION: Activity enhanced by manganese ions and

CC inhibited by p-chloromercuribenzoate.



```
ID SYA MYCPN STANDARD; PRT; 900 AA.
AC P75368;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase [EC 6.1.1.7] (Alanine--tRNA Ligase) (Alars).
GN ALAS OR MPN419 OR MP422.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert R., Plagens H., Pirkel B., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
CC diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A3600041; AAB96070.1; -.
CC F1R; S73748; S73748.
CC HAMAP; MF 00036; -.
CC DR InterPro; IPR002318; tRNA-synt_2c.
CC DR InterPro; IPR006193; tRNA_synt_ala.
CC DR Pfam; PF01411; tRNA-synt_2c; 1.
CC DR PRINTS; PR00980; TRNASYNTHALA.
CC DR TIGRFAMs; TIGR00344; alas; 1.
CC DR PROSITE; PS00860; AA tRNA LIGASE II ALA; 1.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SQ SEQUENCE 900 AA; 103634 MW; 59C1A9FA0140B6E CRC64;

Query Match 4.3%; Score 102.5; DB 1; Length 900;
Best Local Similarity 23.6%; Pred. No. 1.6;
Matches 49; Conservative 32; Mismatches 80; Indels 47; Gaps 7;

QY 3 KIEVINSKHTIKPSTSTELQPKLTLIDLTTPAYPIVFYPIITDHDENLPQTLDL 62
DB 719 KLSSTINSLSHTLNMIANPALNKQKALTQKQ-----NHFHLQVITDL 762
QY 63 R-QALSETTLXYPLSGRVKNLYIDTFEGVPYLEARVNCMDTFLRLRKIECLAEFV 120
DB 763 RKCQALLNELXI-----TVNELKTEDPKWKQQLAEKIKQELLEAKQDKAYVLASFA 815
QY 121 PIKPFM-----EASIDREYPLGVQVNVFD---SGIATGVSVS-----HKLIDGGA 165
DB 816 AVDPFKLLSQVAQAVLNQHNKFLVLLNQNPNNSPMLLQGVSKICQLLKAHFLKGGGS 875
QY 166 DCFKLSKSGAVRGCEENIHPSLSAAL 193
DB 876 NNF-----FRGSPNESVDVSKLQAIL 896

RESULT 8
PRLR_MELGA STANDARD; PRT; 831 AA.
ID PRLR_MELGA
AC Q91094; Q91091; Q91092;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
```

```
DE Prolactin receptor precursor (PRL-R) (TPRLR).
GN PRLR.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.
OX NCBI_TaxID=9103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97057891; PubMed=8902221;
RA Zhou J.F., Zadworny D., Guemene D., Kuhnlein U.;
RT "Molecular cloning, tissue distribution, and expression of the
RT prolactin receptor during various reproductive states in Meleagris
RT gallopavo.";
RL Biol. Reprod. 55:1081-1090 (1996).
RN [2]
RP SEQUENCE OF 82-121 AND 473-522 FROM N.A.
RC TISSUE=Ovary;
RA Pitts G.R., You S.K., Foster D.N., el Halawani M.E.;
RA Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This is a receptor for the anterior pituitary hormone
CC prolactin.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Belongs to the type I cytokine family of receptors.
CC Subfamily 1.
CC -!- SIMILARITY: Contains 4 fibronectin type III domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L76587; AAB01544.1; -.
CC EMBL; U22947; AAA75038.1; -.
CC EMBL; U22924; AAA75039.1; -.
CC HSP; P16471; IIP3.
CC DR InterPro; IPR002996; CR1A.
CC DR InterPro; IPR008957; FN_III-like.
CC DR InterPro; IPR003961; FN_III.
CC DR InterPro; IPR003528; Hemtopoprn_L_F1.
CC Pfam; PF00041; fn3; 4.
CC SMART; SMO0060; FN3; 3.
CC PROSITE; PS01352; HEMATOPO_REC_L_F1; 1.
CC Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
KW SIGNAL 1 23
FT CHAIN 24 831
FT DOMAIN 24 438
FT TRANSVM 439 459
FT DOMAIN 460 831
FT DOMAIN 123 225
FT DOMAIN 228 325
FT DOMAIN 326 428
FT DISULFD 36 46
FT DISULFD 75 86
FT CARBOHYD 59 59
FT CARBOHYD 91 91
FT CARBOHYD 100 100
FT CARBOHYD 112 112
FT CARBOHYD 132 132
FT CARBOHYD 262 262
FT CARBOHYD 303 303
FT CARBOHYD 315 315
FT CARBOHYD 335 335
SQ SEQUENCE 831 AA; 94394 MW; 220916320877FAC1 CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 831;
Best Local Similarity 20.7%; Pred. No. 2.5;
Matches 100; Conservative 57; Mismatches 156; Indels 169; Gaps 21;
```

```

QY 7 SINSKHTIKPSTSTPLQPKYKLTLLDQLTTPPAYVPIVFFYPITDHPML-----POTLADL 62
Db 303 NITKATNEMSGNSDHYVDVTVIVQDDPANVTLEKPKINPKYMLMTWSPPLADV 362
QY 63 RQALSETITLYPLSGRVKNLYIDDFEGVYPLARVNCMDTDFLRKRIKICLNFEVPI 122
Db 363 R---SGWLTLDYEL---RLKP-----EEGEMETVFGV-QOTQY---KMFSLN---PG 402
QY 123 KPF-----SMEAISDERYPPLGVQVNVFDSGIAIGVSVSHKLDGGTADCLX 170
Db 403 KKYIVQIHCKPDHSGNSGSESENYIEIPNDRVKDMIVWIVGLVLSLI-----CLIM 456
QY 171 SWGAVFRGR-----ENIHPSLSEAALL-----PPRRDLPPEKVV 206
Db 457 SWTWLAGIRMTITPLVPQPKIGDTHLETKSRELLSALGCHGFPPTSDEELLI 516
QY 207 DQME-----210
Db 517 EYLEVEDSEDHQLMPSHDSGRPSKNAKITLKTORDSGRSCDSPSLSEKRETCALPS 576
QY 211 ALWF-----AGKK-----VATRFVGVKAISSIODEAKSESVPKPSRVHAVTG 254
Db 577 ALQIQDVRDQAKGAGKGSWESYCVASRE-----KALLFNESAKSSTWPAVQLPNNOPP 631
QY 255 FLWKHLIAASRALTSCTSTSLSTAAQAV-NLRTMRNMETVLDNATGMLFWMAQAILELS 313
Db 632 TFAYHSIVKANKITS---TTTNVVAALVNEERHQSILYSITISGGM---EKQEMENL 687
QY 314 H-----TTPETSDCLKDLVNLINGSVKQNGDYFET-----FKGEGYGRMC 356
Db 688 BSKTQTTTVQVRQNSNEKLPFINAALM-----DYVEVHKVRQDEBPTVLLKHKXSGKIE 743
QY 357 EY 358
Db 744 KY 745

```

## RESULT 9

```

BACB_BACLI STANDARD; PRT; 2607 AA.
ID BACB_BACLI AC 068007;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Bactracin synthetase 2 (BA2) (Includes: ATP-dependent lysine
DE adenylation (LysA) (Lysine activase); ATP-dependent D-ornithine
DE adenylation (D-OrnA) (D-ornithine activase); Ornithine racemase
DE (EC 5.1.1.12)).
GN BACB.
CS Bacillus licheniformis.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_taxid=1402;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RX MEDLINE=98089193; PubMed=9427658;
RA Konz D., Klens A., Schoegendorfer K., Marahiel M.A.;
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
RT 10716: molecular characterization of three multi-modular peptide
RT synthetases."
RL Chem. Biol. 4:927-937(1997).
CC -!- FUNCTION: Activates two amino acids and incorporate a D-ornithine
CC from its second active site into bacitracin.
CC -!- CATALYTIC ACTIVITY: L-ornithine = D-ornithine.
CC -!- COFACTOR: Contains 2 covalently bound phosphopantetheines
CC (potential).
CC -!- PATHWAY: Cyclic peptide antibiotic bacitracin biosynthesis.
CC -!- SUBUNIT: LARGE MULTIZENZYME COMPLEX OF BA1, BA2 AND BA3.
CC -!- DOMAIN: CONSISTS OF TWO MODULES WITH A C-TERMINAL EPIMERIZATION
CC DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO THE PEPTIDE
CC PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR
CC SUBSTRATE ADENYLATION, THIOALATION, CONDENSATION (NOT FOR THE
CC INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND N

```

```

CC METHYLATION (OPTIONAL).
CC MISCELLANEOUS: BACTRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACTRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC -!- SIMILARITY: Contains 2 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF007865; AAC06347.1; -.
CC PIR; T31678; T31678.
CC HSP; P14687; 1AMU.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensatn.
CC InterPro; IPR006163; Pp_bind.
CC InterPro; IPR006162; Ppantne S.
CC Pfam; PF00501; AMP-binding; 2.
CC Pfam; PF00668; Condensation; 3.
CC Pfam; PF00550; PD-binding; 2.
CC PRINTS; PR00154; AMGBINDING
CC PROSITE; PS00012; PHOSPHOPANTHEINE; 1.
CC PROSITE; PS00455; AMP BINDING; 2.
CC PROSITE; PS00075; ACP DOMAIN; 2.
CC KW Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
CC Multifunctional enzyme; Repeat.
CC REPEAT 535 1090
CC FT REPEAT 1547 2141
CC FT DOMAIN 1021 1088
CC FT DOMAIN 2064 2130
CC FT BINDING 1051 1051
CC FT BINDING 2094 2094
CC FT BINDING 2094 2094
CC SQ SEQUENCE 2607 AA; 297474 MW; FF654FAC5B8BBA6F CRC64;

```

```

Query Match 4.2%; Score 99.5; DB 1; Length 2607;
Best Local Similarity 17.6%; Pred. No. 13;
Matches 83; Conservative 72; Mismatches 149; Indels 167; Gaps 19;

```

```

QY 1 MEKIEVINSKHTIK-----PSTSTPLQPKYK-----LTLDDQLTPPAYVPI-VFF 45
Db 985 LDQIPFTINGKADLKALPEPDRRAFAQRYEAPRNCQTEALLSIWQDILPAEQIGINDEF 1044
QY 46 YPITDHFENLPOTLADLQALSETLTL-----72
Db 1045 PDIGCHSLKAFSMAKIQSALAKVEVLKELFNHSTIQDLAAVIAQKQKQVQSDIQAKKK 1104
QY 73 -YYPUSGRVKNLYIDDFEG-----VPYL-----96
Db 1105 EYYPPLSQAQRILYILNQIEEGQTAYNMPFAMKINGELOTKAEKAFRTLKRHESRTSF 1164
QY 97 -----EARNVNCMDTDFLRKLI-----ECLNFEVPIKPFPSMEAISDERYPPLIG---V 140
Db 1165 VTINPEVQVNIIEVTFEMKYRELNCSLRNNQF--IRPFEL-----EKAPLLRAELV 1217
QY 141 QVNVFDSGIAIGVSVSHKLDGGTADCLFKSGAVPFGRCENIHH-----196
Db 1218 RVNAAEHILL--DMHHIISDGVSGILMKWAALY---EEKELAPLKIQYKDYSEWORD 1272
QY 187 -----SUSEAALLFPDRDLDPEKV-----DQEWALWFAKKVATRFVFGVKAISS 233

```

```

RESULT 11
MDR3_CRIGR STANDARD; PRT; 1281 AA.
AC P23174;
ET 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Multidrug resistance protein 3 (P-glycoprotein 3).
GN PGY3 OR PGP3.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae;
OC Cricetulus.
OX NCBI_TaxID=10023;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92135896; PubMed=1685679;
RA Endicott J.A., Sarangi F., Ling V.;
RT *Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
RL gene family".
RL DNA Seq. 2:89-101(1991).
CC -!- FUNCTION: Energy-dependent efflux pump responsible for decreased
CC drug accumulation in multidrug-resistant cells.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: PGP isoforms differ in their drug transport
CC capabilities: PGP1 and PGP2 can mediate MDR, while PGP3 apparently
CC cannot.
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60042; AAA6885.1; -.
CC PIR; I48123; I48123.
CC HSSP; P13569; INED.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR001140; ABC_TM_transpt.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00664; ABC_membrane; 2.
CC Pfam; PF00005; ABC_tran; 2.
CC ProDom; PD000006; ABC_transporter; 2.
CC SMART; SM00382; AAA; 2.
CC PROSITE; P50929; ABC_TM1F; 2.
CC PROSITE; P500211; ABC_TRANSPORTER_1; 1.
CC PROSITE; P50893; ABC_TRANSPORTER_2; 2.
CC ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
CC Multigene family.
KW DOMAIN
FT 1 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT DOMAIN 349 712 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 713 733 POTENTIAL.
FT TRANSMEM 758 778 POTENTIAL.
FT TRANSMEM 834 854 POTENTIAL.
FT TRANSMEM 855 875 POTENTIAL.
FT TRANSMEM 938 958 POTENTIAL.
FT TRANSMEM 975 995 POTENTIAL.
FT DOMAIN 996 1281 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 429 436 ATP (POTENTIAL).
FT NP_BIND 1071 1078 ATP (POTENTIAL).
SQ SEQUENCE 1281 AA; 140866 MW; 2203EF61EBB29602 CRC64;
4.2%; Score 99; DB 1; Length 1281;
Query Match

```

```

Best Local Similarity 22.7%; Pred. No. 5.2;
Matches 89; Conservative 65; Mismatches 156; Indels 82; Gaps 22;
QY 27 KLTLLDQLTPPAYVPIVFFPIIDHDFNLPTQTLADLRQALSETLTLYPLSGRVKNLYI 86
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 KILLDEATSA-----DTSEAEVQAALDKAREGRT-TIVIAHRLS-TVRNADVI 600
QY 87 DPEEGGVYLEARVNCMDMT-----FLLRKIECINEFVPIKPFMSMEALSDER----YPLL 138
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 AGFEDGV-IVROGSHSELQMKGVEFKLVNMQTSQSQILSQEFEVE-LSEKAADGMPN 658
QY 139 GGVQNVVFDSCIAIGVSVS-----EKL-IDGCTADC-----ELK-----SWG 173
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 659 GWKSHIFRNSTKSLKSSRAHHRLLVDADLDANVPVSVFLKVLKLNKTWEPYFVVGTV 718
QY 174 -AVFRGCRENIIHPSLSEAAALFPDRDLPKRYVDQMEALWFAGKKVAT-----RRFVF 226
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 719 CAIVNGALQPAISIIILSEMIATFGPGDDAVKQKCNLPISLVFLGLGVLSFFTFLOGTF 778
QY 227 GVKATISSIQDEAKSESVPKPSR-----VHAVTGFLMKHLIAASRALTSGTSTLSI 278
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 779 G-KASEIITLRLSRMAFKAMLRQDMWFEDYKNSGTALSTRL-ATDRAQVQAGATGRLAL 836
QY 279 AAQ-AVNLRTRNMETVLNDATGNLFWAQAILELSETTPEISDLKLCIDLVNLLNGSVKQ 337
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 837 IAQNTANLGTGIIISFI-----YGMQLTLLLS-VVPFTAVSGIVEM-KMLAGNAKR 886
QY 338 CNGDYFETFKEGYGRMC-EYLDQRTWSSM 368
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 887 DK-----KALEAAGKIATEAENIRTWSL 911

RESULT 12
XAB2 HUMAN
ID XAB2 HUMAN STANDARD; PRT; 855 AA.
AC Q9HCS7; Q9TET6; Q9GHB0; Q9GIW0; Q9NRG6; Q9ULP3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE XPA-binding protein 2 (HCPN protein) (P3898).
GN XAB2 OR HCPN OR KIAA1177.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBUNIT.
RX MEDLINE=20519578; PubMed=10944529;
RA Nakatsu Y., Asatsuma H., Citterio E., Rademakers S., Vermeulen W.,
RA Kamiuchi S., Yeo J.-P., Khaw M.-C., Saijo M., Kodo N., Matsuda T.,
RA Hoeijmakers J.H.J., Tanaka K.;
RT "XAB2, a novel tetratricopeptide repeat protein, involved in
RT transcription-coupled DNA repair and transcription.";
RL J. Biol. Chem. 275:34931-34937(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RA Li Y., Wu T., Xu S., Ren S., Chen Z., Han Z.;
RT "A novel gene expressed in human adrenal gland.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Qin W.X., Zhou X.M., Zhang P.P., Jiang H.Q., Huang Y., Wan D.F.,
RA Su J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A., AND VARIANTS ILE-126; GLN-454 AND THR-702.
RA Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
RA Miyamoto K.B., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

```

RA SEQUENCE FROM N.A.  
 RX TISSUE=Lung;  
 RA MEDLINE=2238257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.R.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.A., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.B.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 [6]  
 RN SEQUENCE OF 28-855 FROM N.A.  
 RC TISSUE=Spleen;  
 RA Chara O., Nagase T., Kikuno R., Okumura K.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 spleen.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 [7]  
 RN SEQUENCE OF 101-855 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20039618; PubMed=10574461;  
 RA Hirose M., Nagase T., Ishikawa K.-I., Kikuno R., Komura N.,  
 RA Chara O.;  
 RT "Characterization of cDNA clones selected by the GeneMark analysis  
 from size-fractionated cDNA libraries from human brain.";  
 RL DNA Res. 6:329-336 (1999).  
 CC -!- FUNCTION: Involved in transcription-coupled repair (TCR) and  
 transcription.  
 CC -!- SUBUNIT: Associates with RNA polymerase II, the TCR-specific  
 proteins CKN1/CSA and ERCC6/CSB, and XPA.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: Contains 11 HAT repeats.  
 CC -!- SIMILARITY: Contains 10 TPR repeats.  
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to  
 framehifts.  
 CC -!- CAUTION: Ref.6 sequence differs from that shown due to  
 framehifts.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AB026111; BAB15807.1; -  
 DR EMBL; AF226051; AAF86951.1; ALT\_FRAME.  
 DR EMBL; AF258567; AAG23770.1; -  
 DR EMBL; AF547265; AANI7847.1; -  
 DR EMBL; BC007208; AAH07208.1; -  
 DR EMBL; BC008778; AAH08778.1; ALT\_INIT.  
 DR EMBL; AK074035; BAB84861.1; ALT\_FRAME.  
 DR EMBL; AB033003; BAA86491.1; -  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0006283; P:transcription-coupled nucleotide-excision r. . .; IDA.  
 DR GO; GO:0006350; P:transcription; IDA.  
 DR InterPro; IPR003107; HAT.

DR InterPro; IPR038940; Prenyl\_trans.  
 DR InterPro; IPR008941; TPR-like.  
 DR InterPro; IPR001440; TPR.  
 DR SMART; SMO0386; HAT; 6.  
 KW DNA repair; Transcription; Nuclear protein; Repeat; TPR repeat;  
 KW Polymorphism.  
 FT REPEAT 15 47 HAT 1.  
 FT REPEAT 33 67 TPR 1.  
 FT REPEAT 48 80 HAT 2.  
 FT REPEAT 90 122 HAT 3.  
 FT REPEAT 110 143 TPR 2.  
 FT REPEAT 124 158 HAT 4.  
 FT REPEAT 146 179 TPR 3.  
 FT REPEAT 160 192 HAT 5.  
 FT REPEAT 256 289 TPR 4.  
 FT REPEAT 270 305 HAT 6.  
 FT REPEAT 355 388 TPR 5.  
 FT REPEAT 369 407 HAT 7.  
 FT REPEAT 395 428 TPR 6.  
 FT REPEAT 433 466 TPR 7.  
 FT REPEAT 484 517 TPR 8.  
 FT REPEAT 498 530 HAT 8.  
 FT REPEAT 519 551 TPR 9.  
 FT REPEAT 532 566 HAT 9.  
 FT REPEAT 554 590 TPR 10.  
 FT REPEAT 571 605 HAT 10.  
 FT REPEAT 679 713 HAT 11.  
 FT VARIANT 126 126 V -> I.  
 FT /FTID=VAR\_016248.  
 FT VARIANT 454 454 R -> Q.  
 FT /FTID=VAR\_016249.  
 FT VARIANT 702 702 A -> T.  
 FT /FTID=VAR\_016250.  
 FT CONFLICT 68 68 Y -> T (IN REF. 2).  
 FT CONFLICT 140 140 L -> M (IN REF. 1).  
 FT CONFLICT 447 447 E -> K (IN REF. 5; AAH08778).  
 FT CONFLICT 680 680 E -> K (IN REF. 2).  
 FT CONFLICT 751 753 SAT -> IP (IN REF. 2).  
 SQ SEQUENCE 855 AA; 100009 MW; CF766917CD65F6FD CRC64;  
 Query Match 4.1%; Score 96.5; DB 1; Length 855;  
 Best Local Similarity 21.1%; Pred.No. 4.7;  
 Matches 70; Conservative 49; Mismatches 114; Indels 99; Gaps 16;  
 QY 167 CFKSWGAVFRGRENIIHPSLS-----BAALLFPDRDLPEKYVDQMEALWFA 215  
 Db 66 CSYKLFYRLKARQAQVHCRCVTDPAEDVNNCHERAFVPMHK--MFLWLDYCCFLMDQ 123  
 QY 216 GKQVATRR-FVFGVKAISIIODE-----AKSESVKPSRVHVTGFLMKHLIAASR 265  
 Db 124 GRVTHRTTFEDRALRALPITQHSRIWPLRLRSLRPLFE-TAVRGYRFRFLKLSPEAAE 182  
 QY 265 ALTSGETSTRSLIAQAQVNLRTMNMETVLDNATGNLFWMAQAILELSTHTTPEISDLKLC 325  
 Db 183 YIEYLKSSDLDEAAQ--RLATVANDERFVSKA-----GKSNYQLWH-----ELC 225  
 QY 326 DLVNLNLSGKVCNSDYFETFKG-----KEGYERM-CEYLDFTQTMSSMEPADYIY---- 375  
 Db 226 DLISQNPDKVQSLNVD--AIIINGLRTFTDQLGKLWCSLADYVIRSGHFKEKARDVVEAI 283  
 QY 376 -----LFSSWTNFFNLDGNGRTSGTWIGVAGKIESASCRFIIIVPTQCGGIEA 424  
 Db 284 RTVMVTVRDFTQVFDISYAQFEESM-----IAAKMETAS-----ELGREED 323  
 QY 425 WVNLE-----EEKMA-----MLRQDDPH 441  
 Db 324 DVDLELRARFEQLISREPLNLNSVILRQNH 355  
 RESULT 13  
 XAB2 MOUSE  
 ID XAB2 MOUSE STANDARD; PRT; 855 AA.  
 AC Q9DCD2; Q8VD75; Q9CVD8;

DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 19-OCT-2003 (Rel. 42, Last annotation update)  
 DE XPA-binding protein 2.  
 GN XAB2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney, and Stomach;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Resole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RL Nature 409:685-690(2001).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.B.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Involved in transcription-coupled repair (TCR) and  
 CC transcription (by similarity).  
 CC -1- SUBUNIT: Associates with RNA polymerase II, the TCR-specific  
 CC proteins CKN1/GSA and ERCC6/CSB, and XPA (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: Contains 11 TAT repeats.  
 CC -1- SIMILARITY: Contains 10 TAT repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AK002890; BAB22435.1; --

DR EXBL; AK008628; BAB25790.1; --  
 DR EXBL; BC021341; AAB21341.1; --  
 DR MGD; MGI.1914689; 0610041014Rik.  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0005515; P:protein binding; ISS.  
 DR GO; GO:0006283; P:transcription-coupled nucleotide-excision r. . .; ISS.  
 DR GO; GO:0006350; P:transcription; ISS.  
 DR InterPro; IPR003107; HAT.  
 DR InterPro; IPR006940; Prenyl trans.  
 DR InterPro; IPR008941; TPR-like.  
 DR InterPro; IPR001440; TPR.  
 DR SMART; SM00386; HAT; 10.  
 KW DNA repair; Transcription; Nuclear protein;  
 KW Repeat; TPR repeat.  
 FT REPEAT 15 47 HAT 1.  
 FT REPEAT 33 67 TPR 1.  
 FT REPEAT 48 80 HAT 2.  
 FT REPEAT 90 122 HAT 3.  
 FT REPEAT 110 143 TPR 2.  
 FT REPEAT 124 158 HAT 4.  
 FT REPEAT 146 179 TPR 3.  
 FT REPEAT 160 182 HAT 5.  
 FT REPEAT 206 289 TPR 4.  
 FT REPEAT 270 305 HAT 6.  
 FT REPEAT 355 388 TPR 5.  
 FT REPEAT 369 407 HAT 7.  
 FT REPEAT 395 428 TPR 6.  
 FT REPEAT 433 466 TPR 7.  
 FT REPEAT 484 517 HAT 8.  
 FT REPEAT 498 530 TPR 8.  
 FT REPEAT 519 551 TPR 9.  
 FT REPEAT 532 566 HAT 9.  
 FT REPEAT 554 590 TPR 10.  
 FT REPEAT 571 605 HAT 10.  
 FT REPEAT 679 713 HAT 11.  
 FT CONFLICT 684 684 A -> T (IN REF. 2).  
 FT CONFLICT 842 842 Q -> L (IN REF. 1; BAB25790).  
 SQ SEQUENCE 855 AA; 59987 MW; 6A5DAGA74E7FCH1D CRC64;  
 Query Match 4.1%; Score 96.5; DB 1; Length 855;  
 Best Local Similarity 21.1%; Pred. NO. 4.7;  
 Matches 70; Conservative 49; Mismatches 114; Indels 99; Gaps 16;  
 QY 167 CFLKSGAVFGCGRENILHPSL-----EALLPFRDLPKDYDQKALWPA 215  
 Db 66 CSYKLYRYLKARAQVGRVTPAYEDVNVNCHERAFVFMHK--MPRLWLDYCOFLMDQ 123  
 QY 216 GKQVATRR-FVFGVKAISSIQDE-----AKSESVKPSRVHVAVTGFLWKHLIAASR 265  
 Db 124 GRVTHTRTFDRALRALPITQHSRIWPLYLRLRSHPLPE-TAVRGYRFLKLPESAE 182  
 QY 266 ALTSGLTSTRLSIAAQVNLRTNMETVLDNATCNLFWAQAILELSHTTPTSDKLC 325  
 Db 183 YIEYKSSDRIDEAQQ--RLATVNVDERFVSKA-----GKSNYQLWH-----ELC 225  
 QY 326 DLVNLGSKVQKCGDYFEFTFKG-----KEGYGRM-CEYLDLDFORTMSSMEAPDIY---- 375  
 Db 226 DLISQNDKVKQSLNVD--AIIRGLTFTDLQSLKWLADYYIRSGHFEKARVISEAI 283  
 QY 376 -----LPSSWTNFFNPLDFGNGRTSWIGVAGKIESASCKFIILVPTQCGSGIEA 424  
 Db 284 RTVMTVRDFTQVDFSAQFESM-----IAAKMETAS-----ELGREED 323  
 QY 425 WVNLE-----EEKWA-----MLEQDPH 441  
 Db 324 DVDELELRLARFEQLISRRP-LNLNSVLLRNPH 355  
 RESULT 14  
 XAB2 RAT  
 ID XAB2 RAT STANDARD; PRT; 855 AA.  
 AC Q99PK0;  
 DT 28-FEB-2003 (Rel. 41, Created)





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:50 ; Search time 13.8278 Seconds  
(without alignments)  
2267.788 Million cell updates/sec

Title: US-09-857-518a-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKLVYGVCHSD.....ERVVKDVRPRFVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	971	55.2	361	T12571	cinnamyl-alcohol d
2	963	54.8	337	S28045	cinnamyl-alcohol d
3	960	54.6	357	T05625	cinnamyl-alcohol d
4	959	54.6	359	S28043	cinnamyl-alcohol d
5	935.5	53.2	354	S72477	probable cinnamyl-
6	881	50.1	356	S71179	cinnamyl-alcohol d
7	881	50.1	360	T08581	cinnamyl-alcohol d
8	853	48.5	363	T05624	cinnamyl-alcohol d
9	819.5	46.6	375	S28044	cinnamyl-alcohol d
10	802	45.6	376	S84604	alcohol dehydrogen
11	740	42.1	357	S39509	alcohol dehydrogen
12	739	42.0	357	S49443	cinnamyl-alcohol d
13	719	40.9	357	S49444	cinnamyl-alcohol d
14	686	39.0	357	S23526	cinnamyl-alcohol d
15	681	38.7	357	S23525	cinnamyl-alcohol d
16	666	37.9	357	T09141	cinnamyl-alcohol d
17	646	36.7	357	T05413	cinnamyl-alcohol d
18	641	36.5	358	S31572	cinnamyl-alcohol d
19	629	35.8	354	S60242	cinnamyl-alcohol d
20	628	35.7	367	T02990	cinnamyl-alcohol d
21	621	35.3	367	T02767	cinnamyl-alcohol d
22	616.5	35.1	362	S45094	cinnamyl-alcohol d
23	613.5	34.9	362	S87125	alcohol dehydrogen
24	609	34.6	355	AC2768	alcohol dehydrogen
25	609	34.6	368	S97548	alcohol dehydrogen
26	602	34.2	341	S31571	cinnamyl-alcohol d
27	599.5	34.1	346	H70860	alcohol dehydrogen
28	599.5	34.1	346	UC1376	alcohol dehydrogen
29	596.5	33.9	349	C69583	alcohol dehydrogen

30	587.5	33.4	355	2	E96751	alcohol dehydrogen
31	581.5	33.1	353	2	D83361	alcohol dehydrogen
32	571.5	32.5	352	2	G82719	alcohol dehydrogen
33	568.5	32.3	349	2	D82563	alcohol dehydrogen
34	566.5	32.2	348	2	AH2824	alcohol dehydrogen
35	561	31.9	348	2	G82645	alcohol dehydrogen
36	555.5	31.6	349	2	E64759	probable alcohol d
37	551.5	31.4	349	2	C90676	alcohol dehydrogen
38	551.5	31.4	349	2	F85526	alcohol dehydrogen
39	548.5	31.2	349	2	H82643	alcohol dehydrogen
40	504	28.7	348	1	H64657	probable cinnamyl-
41	503	28.6	350	2	A71857	zinc-dependent alc
42	475	27.0	358	2	B81302	probable alcohol d
43	474	27.0	365	2	H71808	zinc-dependent alc
44	415.5	23.6	336	1	S76928	probable aryl alco
45	405.5	23.1	241	2	G97602	hypothetical prote

## ALIGNMENTS

### RESULT 1 T12571

C:Species: Mesembryanthemum crystallinum (common ice plant)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 08-Dec-2000  
C:Accession: T12571

R:Michalowski, C.B.; Bohnert, H.J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z17531

A:Accession: T12571

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-361 <MIC>

A:Cross-references: EMBL:U79770; NID:G1724109; PID:G1724110

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: alcohol metabolism; oxidoreductase; zinc

F:32-340/Domain: long-chain alcohol dehydrogenase homology <LAD>  
F:51,73,167/Binding site: zing, catalytic (Cys, His, Cys) #status predicted

Query Match 55.2%; Score 971; DB 2; Length 361;

Best Local Similarity 55.4%; Pred. No. 3e-73;

Matches 190; Conservative 52; Mismatches 63; Indels 38; Gaps 6;

Qy	2	TGATDVRFKLVYGVCHSDHMAKNDWGTSTYPIVPGHELIVGVTVGVCKVKFK-SWRQ	60
Db	36	TGEQDVTFRKLVYGVCHSDHLYIKNEGNNAVYPAIPGHEIVGVTVGVCKVKFKVDKV	95
Qy	61	GRCLHGLRPTCENCILHLENVCPNLIQTYGSKYVYDGTMTYGVYNNMVTDEHFIYRIP	120
Db	96	GVGCMVGSQR-SCSCENHLENVCPNLIYGYSTYYDGLTYGYSYDLMVVEHFVAVRIP	154
Qy	121	DNPLDGAAPLLCAGITYSPWYYGLDKPMHMGVEMPRFRPRPPLNPLGLWGRSLQS	180
Db	155	DNMALDATAPLLCAGVTYSPKHFELDKDGLHGV-----VGLGG-----	195
Qy	181	LVPLLEKGSYG-----TSPALHGS-----LRTDQDQMEANSTMDGIDT	223
Db	196	LGNMVFGRKAFKAVTVISTSPNKKDDEAVNRLGADSFVVSREPEQMSAMGLDIDT	255
Qy	224	VPVPRPLETISLITKNGKVTVGVIAVQPLDLPVPLLIIGRQWVAGSAIGMKETOEMID	283
Db	256	VSAHPPLPLLLGLKSKGMIMVGVDPKPLEPFPPLQGRKILANGSCIGMKETOEMID	315
Qy	284	FAEHNTADIEVIPIDYLNAMERVVKDVRFRFVIDVENTL	326
Db	316	FAAKHDIKSDIEVPMYDVTAMERLLKGDVRYFVIDVANTL	358

### RESULT 2 S28045

C:Species: Petroselinum Crispum (parsley)  
Cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3 - parsley (fragment)

C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
 C/Accession: S28045  
 R/Kiedrowski, S.; Kwallack, P.; Hahlbrock, K.; Somssich, I.E.; Dangl, J.L.  
 EMBO J. 11, 4677-4684, 1992  
 A/Title: Rapid activation of a novel plant defense gene is strictly dependent on the Ara  
 A/Reference number: S28043; MUID:93099840; PMID:1464303  
 A/Accession: S28045  
 A/Molecule type: mRNA  
 A/Residues: 1-337 <KIE>  
 A/Cross-references: EMBL:X67817; NID:g20446; PIDN:CAA48028.1; PID:g836638  
 A/Note: the authors did not translate the codon for residue 337  
 C/Genetics:  
 A/Genes: ELI3  
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C/Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc  
 F:12-320/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:27,49,143/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 54.8%; Score 963; DB 2; Length 337;  
 Best Local Similarity 55.2%; Pred. No. 1.3e-72;  
 Matches 191; Conservative 46; Mismatches 65; Indels 44; Gaps 5;

QY 2 TGATDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHVLGVVTVGVCKVKKFK-----S6  
 Db 12 TGDNDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHVLGVVTVGVCKVKKFK-----S6  
 QY 57 --SWRGRCWLEHRLPTCENCILHLENYCPNLIQTYGSKYVDGTYGYSNNMVTDEH 114  
 Db 72 GVGCLVGSCL-----SCNCDSDSNNCAKQVYAFYVDSGTYGYADSVADQH 124  
 QY 115 FIVRPEMLPLDGAAPLLCAGITTYSPWRYVGLDKPGHMLGVWPRFRSRPPLNLPGLW 174  
 Db 125 FIVRPEMLPLDGAAPLLCAGITTYSPWRYVGLDKPGHMLGVWPRFRSRPPLNLPGLW 174  
 QY 175 GSRLQSLVPLKKEGSGYTSALMHS-----LIRTDQDQMEAMSTMDGI 220  
 Db 171 G--LGHVAVKAKAFGANVTVISTESKKQALEKLGADFLVSSDSQMQATGFLHGI 228  
 QY 221 IDTVPAVRPLEPLISLLKTNKGVTVGIAVQPLDLPVFPPLIIGRKMVAGSAIGMKETQE 280  
 Db 229 IDTVSALHPVPLGLLKVNGKLVWVGAPEKLEPLFPVPLLMGRKVLASNGIGLKEQ 288  
 QY 281 MIDFAEHNITADIEVPIPIYDYLNTAMERVKVKKVDFRFRFVIDVENTL 326  
 Db 289 MIDFAAQNITADIEVPIYDYLNTAMERLVKSDVRYRFFVIDVANTI 334

RESULT 3  
 T05625  
 cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3-1 - Arabidopsis thaliana  
 N/Alternate names: protein F20D10.100  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 15-Oct-1999  
 C/Accession: S28044  
 R/Bevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.  
 submitted to the Protein Sequence Database, February 1999  
 A/Reference number: Z15420  
 A/Cross-references: EMBL:X67816; NID:g16266; PIDN:CAA48027.1; PID:g16267  
 A/Molecule type: DNA  
 A/Residues: 1-357 <BEV>  
 A/Cross-references: EMBL:AL035538  
 A/Experimental source: cultivar Columbia; BAC clone F20D10  
 R/Kiedrowski, S.; Kwallack, P.; Hahlbrock, K.; Somssich, I.E.; Dangl, J.L.  
 EMBO J. 11, 4677-4684, 1992  
 A/Title: Rapid activation of a novel plant defense gene is strictly dependent on the Ara  
 A/Reference number: S28043; MUID:93099840; PMID:1464303  
 A/Accession: S28044  
 A/Molecule type: mRNA  
 A/Residues: 1-5, 'Q', 7-16, 'N', 18-19, 'V', 21-357 <KIE>  
 A/Cross-references: EMBL:X67816; NID:g16266; PIDN:CAA48027.1; PID:g16267  
 C/Genetics:  
 A/Genes: ELI3-1  
 A/Map position: 4

A/Introns: 29/2; 67/2; 290/1

A/Note: F20D10.100  
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C/Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc  
 F:31-339/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:46,68,162/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 54.6%; Score 960; DB 2; Length 357;  
 Best Local Similarity 55.4%; Pred. No. 2.4e-72;  
 Matches 190; Conservative 47; Mismatches 68; Indels 38; Gaps 5;

QY 2 TGATDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHVLGVVTVGVCKVKKFKSM-RQ 60  
 Db 31 TGEKDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHVLGVVTVGVCKVKKFKSM-RQ 60  
 QY 61 GRCLHGRRLPTCENCILHLENYCPNLIQTYGSKYVDGTYGYSNNMVTDEHFIIRIP 120  
 Db 91 GVGIMAGSCR-SCDSCNDSDENYCPNLIQTYGSKYVDGTYGYSNNMVTDEHFIIRIP 149  
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYVGLDKPGHMLGVWPRFRSRPPLNLPGLWSRLQS 180  
 Db 150 DNLPLDGAAPLLCAGITTYSPWRYVGLDKPGHMLGVWPRFRSRPPLNLPGLWSRLQS 190  
 QY 181 LVPPLIKEGSGYTSALMHS-----LIRTDQDQMEAMSTMDGIIDT 223  
 Db 191 LGHVAVKAKAFGANVTVISTESKKQALEKLGADFLVSRDPKQMKDAMTMDGIIDT 250  
 QY 224 VPAVRPLEPLISLLKTNKGVTVGIAVQPLDLPVFPPLIIGRKMVAGSAIGMKETQEMID 283  
 Db 251 VSATHEPLPLGLLKVNGKLVWVGAPEKLEPLFPVPLLMGRKVLASNGIGLKEQEMVT 313  
 QY 284 FAEHNITADIEVPIPIYDYLNTAMERVKVKKVDFRFRFVIDVENTL 326  
 Db 311 LAGKNITADIEVPIYDYLNTAMERLVKSDVRYRFFVIDVANTM 353

#### RESULT 4

S28043  
 cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3-2 - Arabidopsis thaliana  
 N/Alternate names: protein F20D10.110  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 15-Oct-1999  
 C/Accession: S28043; T05626  
 R/Kiedrowski, S.; Kwallack, P.; Hahlbrock, K.; Somssich, I.E.; Dangl, J.L.  
 EMBO J. 11, 4677-4684, 1992  
 A/Title: Rapid activation of a novel plant defense gene is strictly dependent on the Ara  
 A/Reference number: S28043; MUID:93099840; PMID:1464303  
 A/Accession: S28043  
 A/Molecule type: mRNA  
 A/Residues: 1-359 <KIE>  
 A/Cross-references: EMBL:X67815; NID:g16268; PIDN:CAA48026.1; PID:g16269  
 R/Bevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.  
 submitted to the Protein Sequence Database, February 1999  
 A/Reference number: Z15420  
 A/Cross-references: EMBL:AL035538  
 A/Molecule type: DNA  
 A/Residues: 1-359 <BEV>  
 A/Cross-references: EMBL:AL035538  
 A/Experimental source: cultivar Columbia; BAC clone F20D10  
 C/Genetics:  
 A/Genes: ELI3-2  
 A/Map position: 4  
 A/Introns: 29/2; 67/2; 290/1  
 A/Note: F20D10.110  
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C/Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc  
 F:31-339/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:46,68,162/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 54.6%; Score 959; DB 1; Length 359;  
 Best Local Similarity 55.1%; Pred. No. 3e-72;  
 Matches 193; Conservative 39; Mismatches 68; Indels 50; Gaps 5;



A:Map position: 4  
 A:Introns: 33/2; 71/2; 242/3; 294/1  
 A:Note: F22F8.230  
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc  
 F:31-339/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F:186-215/Region: beta-alpha-beta NADP nucleotide-binding fold  
 F:50,72,166/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted  
 F:103,106,109,117/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 50.1%; Score 881; DB 2; Length 360;  
 Best Local Similarity 49.1%; Pred. No. 9.5e-86;  
 Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGATDVRKVLVCGVCHSDIHMKNKDWGTSYPIVPGHVLGVVTEVGVCKVKKFSWRQ 60  
 DB 34 DNGENDVTVKILFCGVCHSDIHMKNKDWGTSYPIVPGHVLGVVTEVGVCKVKKFSWRQ 93

QY 61 -----GRCWLHGRRLPTCENCIHLENYCPNLQIOTYGSKYVDGTMVYGGYNNMTDE 113  
 DB 94 VGVGVLISSC-----QSCSCDQDLENYCPQMSFTYNAIGSDGKYNKYGSYSENIVDQ 146

QY 114 HFIIVIPDMLPDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPRLNPLGL 173  
 DB 147 RFVLRPENLPDSGAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPRLNPLGL 192

QY 174 WSRQLQSLVPLPIKEGSGYTSALMHS-----LLRTDQDQMEAMST 216  
 DB 193 GG-----LGHVAVKIGKAFGLKVTYSSSTKAEAINHLGADSPFLVTDPOQKAAIGT 247

QY 217 MDGIIDTVPAPVPLEPLISLLKTKNGVTVVGVIAVQPLDLPVPLIIGRKNWAGSAIGMK 276  
 DB 248 MDYIIDTISAVHALYPLGLKLVKNGKLIAGLPEKLEPMPLVLGRKNWAGSDVGGMK 307

QY 277 ETQEMIDFAEHNITADIEVIPDYINTAMERVVKKVDFRFRVIDVENTL 326  
 DB 308 ETQEMIDFAEHNITADIEVIPDYINTAMERVVKKVDFRFRVIDVANS 357

RESULT 8  
 T05624  
 cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) LCADa - Arabidopsis thaliana  
 N:Alternate names: protein F20D10.90  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jan-2000  
 C:Accession: T05624  
 R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.H.  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15420  
 A:Accession: T05624  
 A:Molecule type: DNA  
 A:Residues: 1-363 <BEV>  
 A:Cross-references: EMBL:AL035538  
 A:Experimental source: cultivar Columbia; BAC clone F20D10  
 C:Genetics:  
 A:Gene: LCADa  
 A:Map position: 4  
 A:Introns: 34/2; 72/2; 169/1; 243/3; 295/1  
 A:Note: F20D10.90  
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc  
 F:32-340/Domain: long-chain alcohol dehydrogenase homology <LAD>  
 F:187-216/Region: beta-alpha-beta NADP nucleotide-binding fold  
 F:51,73,167/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted  
 F:104,107,110,118/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 48.5%; Score 853; DB 2; Length 363;  
 Best Local Similarity 47.9%; Pred. No. 2.1e-63;  
 Matches 171; Conservative 50; Mismatches 72; Indels 64; Gaps 5;

QY 1 ETGATDVRKVLVCGVCHSDIHMKNKDWGTSYPIVPGHVLGVVTEVGVCKVKKFSWRQ 60  
 DB 35 KTGBEVRKVLVCGVCHSDIHMKNKDWGTSYPIVPGHVLGVVTEVGVCKVKKFSWRQ 94

QY 57 -----SWRQRCMLHGLRLPTCENCIHLENYCPNLQIOTYGSKYVDGTMVYGGYNN 108  
 DB 95 VGVGVLISSC-----TCSCREDQENYCTKAIATYNGVHHDGTINNGYSDH 142

QY 109 MVTDEHFIIVIPDMLPDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPRLNPLGL 168  
 DB 143 IVDYERAVKIPHTLPVSNAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPRLNPLGL 191

QY 169 NLPGLW-----GSRQLQSLVPLPIKEGSGYTSALMHS-----LLRTDQDQ 209  
 DB 192 GLGGLGHIGVRFPAK-----AFGKVTYVSSSTGKSKALDNLGADGFLVSTDEQ 241

QY 210 MEAAMSTMDGIIDTVPAPVPLEPLISLLKTKNGVTVVGVIAVQPLDLPVPLIIGRKNWAG 269  
 DB 242 MKAAMSTMDGIIDTVPAPVPLEPLISLLKTKNGVTVVGVIAVQPLDLPVPLIIGRKNWAG 301

QY 270 SAIGGMKETOEMIDFAEHNITADIEVIPDYINTAMERVVKKVDFRFRVIDVENTL 326  
 DB 302 SGIGGMKETOEMIDFAEHNITADIEVIPDYINTAMERVVKKVDFRFRVIDVENTL 358

RESULT 9  
 D84606  
 alcohol dehydrogenase (EC 1.1.1.1) [similarity] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Aug-2002  
 C:Accession: D84606  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617157  
 A:Accession: D84606  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-375 <STO>  
 A:Cross-references: GB:AE002093; MID:g4417281; PIDN:AAD20406.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2G21890  
 A:Map position: 2  
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C:Keywords: oxidoreductase; zinc  
 F:44,66,160/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 46.6%; Score 819.5; DB 2; Length 375;  
 Best Local Similarity 47.6%; Pred. No. 1.3e-60;  
 Matches 167; Conservative 48; Mismatches 85; Indels 51; Gaps 6;

QY 1 ETGATDVRKVLVCGVCHSDIHMKNKDWGTSYPIVPGHVLGVVTEVGVCKVKKFSWRQ 60  
 DB 28 ENGENDVTVKILFCGVCHSDIHMKNKDWGTSYPIVPGHVLGVVTEVGVCKVKKFSWRQ 87

QY 61 -----GRCWLHGRRLPTCENCIHLENYCPNLQIOTYGSKYVDGTMVYGGYNNMTDE 113  
 DB 88 VGVGVLISSC-----QSCSCDQDLENYCPQMSFTYNAIGSDGKYNKYGSYSENIVDQ 140

QY 114 HFIIVIPDMLPDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPRLNPLGL 172  
 DB 141 RFVLSPIDPLSDGAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPRLNPLGL 186

QY 173 WSRQLQSLVPLPIKEGSGYTSALMHS-----LLRTDQDQMEAMST 215  
 DB 187 LGG-----LGHVAVKIGKAFGLKVTYSSSTKAEAINHLGADSPFLVTDPOQKAAIGT 241

QY 216 TMDGIIDTVPAPVPLEPLISLLKTKNGVTVVGVIAVQPLDLPVPLIIGRKNWAGSAIGMK 275  
 DB 242 TMDGIIDTVPAPVPLEPLISLLKTKNGVTVVGVIAVQPLDLPVPLIIGRKNWAGSAIGMK 301

QY 276 KETOEMIDFAEHNITADIEVIPDYINTAMERVVKKVDFRFRVIDVENTL 326  
 DB 302 KETOEMIDFAEHNITADIEVIPDYINTAMERVVKKVDFRFRVIDVENTL 352

```
RESULT 10
E84604
alcohol dehydrogenase (EC 1.-.-.-) [similarity] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Aug-2002
C:Accession: E84604
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanRaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84604
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <STO>
A:Cross-references: GB:A802093; NID:g4417268; PIDN:AAD20393.1; GSPDB:GNO0139
C:Genetics:
A:Gene: At2g21730
A:Map position: 2
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase; zinc
F:44,66,161/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 45.6%; Score 802; DB 2; Length 376;
Best Local Similarity 47.2%; Pred. No. 3.9e-59;
Matches 166; Conservative 47; Mismatches 87; Indels 52; Gaps 7;

QY 1 ETGATDVRFKVLGCVCHSDIHMAKNDWGTSTYPIVPGHELVGVVTEYGVCKVKFKSWRQ 60
DB 28 ENGENDVTVKLFCGVCHSDLHTIKNHGFRYPIIPGHEIVGATGKVKWTFKQGR 87
QY 61 -----GRCWLHGRLEPTCENCIIHLENYCPNLIQTYGSKYYDGT-MTYGYSNNMVT 112
DB 88 VGVGVIIIGSC-----QSCSCNQDLENYCPKVFTYNSRSSDGTSRNQGYSDVIIVD 140
QY 113 EHFIVRPDNLPLGAPLLCAGITTYSPWRYGLDK-PGMHLGVWPRFRFRPLNLP 171
DB 141 HRFVLSIPDGLPSGAPLLCAGITTYSPMKYGYMGTKESGRKLVN-----186
QY 172 GLWGRSLQSLVPELIKESGYSTSPALMHS-----LLRTDQDQMEAM 214
DB 187 GLGG-----LGHIAVKIGKAGLRVTVISRSEREREALDGLADSFVITDSQCKEAV 241
QY 215 STMDGLDTPVAVRPLEPLISLLKXNGKVTVGVIAVQPLDLVPVPLIIGRKNVAGSITGG 274
DB 242 GTMDFIIDTVAEHALPLFLSKLVKNGKLVAGLPEKPLDLPFSLVLRKMWGSGQIGG 301
QY 275 MKETQEMIDFAAENITADIEVPIIDYLTAMERVVKDVRFRFVIDVENTL 326
DB 302 MKETQEMLEFCAKHIVSDIBLIKMSDINSAMDRLAKSDVRYFRFVIDVANSL 353

RESULT 11
S39509
cinamyl-alcohol dehydrogenase (EC 1.1.1.195) - Norway spruce
C:Species: Picea abies (Norway spruce)
C>Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 15-Oct-1999
C:Accession: S39509
R:Galliano, H.; Cabane, M.; Eckerskorn, C.; Lottespeich, F.; Sandermann Jr., H.; Ernst, D.
Plant Mol. Biol. 23, 145-156, 1993
A:Title: Molecular cloning, sequence analysis and elicitor-/ozone-induced accumulation o
A:Reference number: S39509; MUID:94033286; PMID:8219046
A:Accession: S39509
A:Molecule type: mRNA
A:Residues: 1-357 <GAL>
A:Cross-references: EMBL:X72675; NID:g393442; PIDN:CAA51226.1; PID:g393443
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
```

```
Query Match 42.1%; Score 740; DB 2; Length 357;
Best Local Similarity 42.7%; Pred. No. 5.4e-54;
Matches 147; Conservative 61; Mismatches 86; Indels 50; Gaps 5;

QY 3 GATDVRFKVLGCVCHSDIHMAKNDWGTSTYPIVPGHELVGVVTEYGVCKVKFKSWRQ-- 60
DB 33 GPEDVIVKVIYCGICHSDLVQMRNEMGMSNYPMVPGHEVGVVTEIGSEVKKFKVGEHV 92
QY 61 -----GRCWLHGRLEPTCENCIIHLENYCPNLIQTYGSKYYDGTMTYGGYSNNMVTDEHF 115
DB 93 VGVGVIIIGSC-----RSCNCNCSMEQYCSKEIWTYNDVNHDTPTQGGFASMVVDQMF 145
QY 116 IVRIPDNLPLGAPLLCAGITTYSPWRYGLDKPGMHLGVWPRFRFRPLNLPGLMW 175
DB 146 VVRIPENLPLEQAAPLLCAGVTYVSPMKFGMTBPKKCGI-----LGLGVGH 194
QY 176 SRLQSLVPLLIKESGYSTSPALMHS-----LLRTDQDQMEAMSTWD 218
DB 195 MG-----VKIAKAFGLHVTVISSDKKKEALEVLGADAYLVSKDAKQKQEAARSJD 246
QY 219 GIIDTPVAVRPLEPLISLLKXNGKVTVGVIAVQPLDLFPFLLIIGRKNVAGSITGGMET 278
DB 247 YIMDTIPVAHPLPYALLKXNGKLVMLGVVPEPLHFVTELLILGRSITAGSFIGSMET 306
QY 279 QEMIDFAAENITADIEVPIIDYLTAMERVVKDVRFRFVIDV 322
DB 307 QETLDFCAEKVKYSMIEVVGGLDYINTAMERLVKNDVRYRFVVDV 350

RESULT 12
S49443
cinamyl-alcohol dehydrogenase (EC 1.1.1.195) A - loblolly pine
C:Species: Pinus taeda (loblolly pine)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Dec-2000
C:Accession: S49443; S55285; S55284
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.; O'Malley, D.
submitted to the EMBL Data Library, September 1994
A:Description: Genetic analysis of Cad in loblolly pine. Single gene inheritance, molec
A:Reference number: S49443
A:Accession: S49443
A:Molecule type: mRNA
A:Residues: 1-357 <MAC>
A:Cross-references: EMBL:Z37991; NID:9558384; PIDN:CAA6072.1; PID:g558385
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.R.; O'Malley, D.M.
Mol. Gen. Genet. 247, 537-545, 1995
A:Title: Genetic analysis of cinamyl alcohol dehydrogenase in loblolly pine: single ge
A:Reference number: S55285; MUID:95327049; PMID:7603432
A:Accession: S55285
A:Molecule type: mRNA
A:Residues: 63-123;189-333-357 <MAW>
A:Cross-references: EMBL:Z37991
C:Genetics:
A:Introns: 123/3; 198/3
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 42.0%; Score 739; DB 2; Length 357;
Best Local Similarity 43.5%; Pred. No. 6.5e-54;
Matches 147; Conservative 63; Mismatches 90; Indels 38; Gaps 5;

QY 3 GATDVRFKVLGCVCHSDIHMAKNDWGTSTYPIVPGHELVGVVTEYGVCKVKFKSWRQ-- 60
DB 33 GPEDVIVKVIYCGICHSDLVQMRNEMGMSHYPMVPGHEVGVVTEIGSEVKKFKVGEHV 92
QY 61 -----GRCWLHGRLEPTCENCIIHLENYCPNLIQTYGSKYYDGTMTYGGYSNNMVTDEHF 115
DB 93 VGVGVIIIGSC-----RSCNCNCSMEQYCSKEIWTYNDVNHDTPTQGGFASMVVDQMF 145
QY 116 IVRIPDNLPLGAPLLCAGITTYSPWRYGLDKPGMHLGVWPRFRFRPLNLPGLMW 164
```

```
Db 146 VVRIPENLPLEQAAPLLCAGVTFS PMKHFAVTEPKKCGILGLGVGHGKIAKAF-- 203
Qy 165 RPPLNPLCGLWSRLQSLVPLPKEGSGVTSPALMHSILRTDQDMRAAMSTMGIIDTV 224
Db 204 --GLHVTVISSDKKKEAMEVLGADAT-----LVSKDTEKMERASLDYIMDTI 252
Qy 225 PAVRPLEPLISLLKNTGKVVTVGIAVQDLDPVPLIIGRKQWAGSAIGGMKETQEMIDF 284
Db 253 PVAHPLEPLALLKNTGKVLMLGVVPEPLHFTVPLPLIGRSIAGSPFGGMEETQETLDF 312
Qy 285 AAENHNTADIEVIPIDYLTAMERVVKKDVRFRFVIDV 322
Db 313 CAEKKVSSMIEVGLDYINTAMERLEKNDVRYRFVVDV 350

RESULT 13
S49444
C:cinamyl-alcohol dehydrogenase (EC 1.1.1.195) B - loblolly pine
C:Species: Pinus taeda (loblolly pine)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Oct-1999
C:Accession: S49444; S5286; S5295
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.; O'Malley, D.
submitted to the EMBL Data Library, September 1994
A:Description: Genetic analysis of Cad in loblolly pine. Single gene inheritance, molecu
A:Reference number: S49443
A:Accession: S49444
A:Molecule type: mRNA
A:Residues: 1-357 <MAC>
A:Cross-references: EMBL:Z37992; NID:9553386; PIDN:CAA86073.1; PID:9553387
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.; O'Malley, D.M.
Mol. Gen. Genet. 247, 537-545, 1995
A:Title: Genetic analysis of cinamyl alcohol dehydrogenase in loblolly pine: single gen
A:Reference number: S55285; MUID:95327049; PMID:7603432
A:Accession: S55286
A:Molecule type: mRNA
A:Residues: 63-123; 189-198; 333-334, 'E', 336-357 <MAW>
A:Cross-references: EMBL:Z37992
C:Genetics:
A:Introns: 123/3; 198/3
A:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 40.9%; Score 719; DB 2; Length 357;
Best Local Similarity 41.6%; Pred. No. 3e-52;
Matches 143; Conservative 63; Mismatches 88; Indels 50; Gaps 5;

Qy 3 GATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVTVTEVGCKVKPKSWRQ-- 60
Db 33 GPEDVIVKVIYCGICHSDLVQKNGMNSHYPVPGHEVVGIVTEIGSRVKKPKVGEHV 92
Qy 61 -----GRCLMHLGRLPTECNCIHLENYCPNLQIOTYGSKYVDGTMVGGYNNMVTDEHF 115
Db 93 VGCIVGSC-----RSCGNCNSMEQYCSKRTWYNDVNDHGTPTCGFASSVVDQMF 145
Qy 116 IVRIPNLPDGAAPLLCAGITTYSPWRYGLDKPGMELGVEMFRFRSRPPLNPLGLMG 175
Db 146 VVRIPENLPLEQAAPLLCAGVTFS PMKHFAVTEPKKCGI-----LGLGGVGH 194
Qy 176 SRLQSLVPLPKEGSGVTS PALMHS-----LLRTDQDMRAAMSTM 218
Db 195 MG-----VKIAKAFGLHVTVISSDKKKEAMEVLGADAYLVSKDTEKMEAEBSLD 246
Qy 219 GIIDTVPAVRPLEPLISLLKNTGKVVTVGIAVQDLDPVPLIIGRKQWAGSAIGGMKET 278
Db 247 YIMDTIPVAHPLEPLALLKNTGKVLMLGVVPEPLHFTVPLPLIGRSIAGSPFGGMEET 306
Qy 279 QEMIDFAAENHNTADIEVIPIDYLTAMERVVKKDVRFRFVIDV 322
Db 307 QETLDFCAEKKVSSMIEVGLDYINTAMERLEKNDVRYRFVVDV 350
```

```
RESULT 14
S23526
C:cinamyl-alcohol dehydrogenase (EC 1.1.1.195) CAD19 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C:Accession: S23526
R:Knight, M.E.; Halpin, C.; Schuch, W.
Plant Mol. Biol. 19, 793-801, 1992
A:Title: Identification and characterisation of cDNA clones encoding cinamyl alcohol de
A:Reference number: S23525; MUID:92353388; PMID:1643282
A:Accession: S23526
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-357 <KNI>
A:Cross-references: EMBL:X62344; NID:919840; PIDN:CAA44217.1; PID:919841
C:Genetics:
A:Gene: CAD19
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 39.0%; Score 686; DB 2; Length 357;
Best Local Similarity 43.1%; Pred. No. 1.7e-49;
Matches 146; Conservative 56; Mismatches 99; Indels 38; Gaps 8;

Qy 2 TGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVTVTEVGCKVKPKSWRQ 61
Db 32 TGPEDVQVKVLYCGLCHSDLHQVNDLGMNYPILVPGHEVVGKRVGEVADVSKFX---VG 88
Qy 62 RCMHLGRLPTECEN---CIHLENYCPNLQIOTYGSKYVDGTMVGGYNNMVTDEHFIVR 118
Db 89 DTGVGGLLVGSCRCNGCPCKREIEQYCNKKIWNQNDVTDGKPTCGGFANSVVDQNFVVK 148
Qy 119 IPNLPDGAAPLLCAGITTYSPWRYGLDKPGMELGVEMFRFRSRPPLNPLGLMSRL 178
Db 149 IPEGMAPEQAAPLLCAGITTYSPFNHFGFNQSGPRGGI-----LGLGGVGHMGV 197
Qy 179 QSLVPLPKEGSGVET-----SPALMH-----SLRTDQDMRAAMSTMGIIDTV 224
Db 198 K-----TAKAMGHVTVISSNKKRQEALEHLGADYLVSSDDTKMQEADSLYIIDTV 252
Qy 225 PAVRPLEPLISLLKNTGKVVTVGIAVQDLDPVPLI-IGRKWAGSAIGGMKETQEMID 283
Db 253 PVGHELELXLKIDKLLIGVINTPLQF-ISPVMVLGRKSTGFIGSMKETEMLD 311
Qy 284 FAENHNTADIEVIPIDYLTAMERVVKKDVRFRFVIDV 322
Db 312 FCXKEGVTSQIEIVKMDYINTAMERLEKNDVRYRFVVDV 350

RESULT 15
S23525
C:cinamyl-alcohol dehydrogenase (EC 1.1.1.195) CAD14 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C:Accession: S23525
R:Knight, M.E.; Halpin, C.; Schuch, W.
Plant Mol. Biol. 19, 793-801, 1992
A:Title: Identification and characterisation of cDNA clones encoding cinamyl alcohol de
A:Reference number: S23525; MUID:92353388; PMID:1643282
A:Accession: S23525
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-357 <KNI>
A:Cross-references: EMBL:X62343; NID:919838; PIDN:CAA44216.1; PID:919839
C:Genetics:
A:Gene: CAD14
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:14 ; Search time 44.1542 Seconds  
(without alignments)  
3229.910 Million cell updates/sec

Title: US-09-857-518A-6  
Perfect score: 2373  
Sequence: 1 MEKIEVINSKTIKPTSTSS.....MAMLEQDPFLALASPKTLLI 452

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 25:\*
- 1: sp archaea:\*
  - 2: sp bacteria:\*
  - 3: sp fungi:\*
  - 4: sp human:\*
  - 5: sp invertebrate:\*
  - 6: sp mammal:\*
  - 7: sp mhc:\*
  - 8: sp organelle:\*
  - 9: sp phase:\*
  - 10: sp plant:\*
  - 11: sp rodent:\*
  - 12: sp virus:\*
  - 13: sp vertebrate:\*
  - 14: sp unclassified:\*
  - 15: sp rvirus:\*
  - 16: sp bacteriap:\*
  - 17: sp archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2373	100.0	452	10 Q9FVF1	Q9fvl1 fragaria an
2	2043	86.1	455	10 Q8GTM5	Q8gtm5 fragaria ve
3	538.5	22.7	455	10 Q9L088	Q9l088 arabidopsis
4	538	22.7	456	10 Q8GV04	Q8gv04 capsicum ch
5	506	21.3	426	10 Q9FGV4	Q9fgv4 arabidopsis
6	501	21.1	443	10 Q9FL40	Q9fl40 arabidopsis
7	498.5	21.0	443	10 Q9LI71	Q9li71 arabidopsis
8	498	21.0	436	10 Q9FTM1	Q9ftm1 arabidopsis
9	485	20.4	446	10 Q23392	Q23392 arabidopsis
10	468.5	19.7	435	10 Q9FYM0	Q9fym0 arabidopsis
11	454.5	19.2	428	10 Q9FLW4	Q9flw4 arabidopsis
12	450	19.0	99	10 Q23943	Q23943 fragaria ve
13	448	18.9	435	10 Q23393	Q23393 arabidopsis
14	438	18.5	474	10 Q94F74	Q94ft4 papaver som
15	406	17.1	443	10 Q8GZU0	Q8gz0 catharantha
16	404.5	17.0	433	10 Q64988	Q64988 clarkia bre

17	403.5	17.0	431	10 Q9SQ00	Q9sq00 clarkia con
18	398.5	16.8	433	10 Q9SPU3	Q9spu3 clarkia bre
19	390.5	16.5	431	10 Q9SQ01	Q9sq01 clarkia con
20	387.5	16.3	439	10 Q9ZTK5	Q9ztk5 catharanth
21	384	16.2	430	10 Q9SQ02	Q9sq02 clarkia con
22	318	13.4	328	10 Q9SPZ9	Q9spz9 clarkia con
23	285	12.0	219	10 Q9M5K9	Q9m5k9 euphorbia e
24	262	11.0	430	10 Q9SMW7	Q9smw7 arabidopsis
25	259.5	10.9	449	10 Q7XPK7	Q7xpk7 oryza sativ
26	253	10.7	426	10 Q8LE88	Q8le88 arabidopsis
27	253	10.7	426	10 Q9FFQ7	Q9ffq7 arabidopsis
28	248	10.5	439	10 Q84QG6	Q84qg6 hordeum vul
29	245.5	10.3	435	10 Q8GSM7	Q8gsm7 nicotiana t
30	242	10.2	446	10 Q23917	Q23917 dianthus ca
31	242	10.2	461	10 Q9LE83	Q9le83 arabidopsis
32	234.5	9.9	461	10 Q9LF70	Q9lf70 arabidopsis
33	233.5	9.8	445	10 Q23918	Q23918 dianthus ca
34	232	9.8	431	10 Q9SST8	Q9sst8 ipomoea bat
35	232	9.8	442	10 Q23916	Q23916 dianthus ca
36	231	9.7	424	10 Q9FTG9	Q9ftg9 oryza sativ
37	230	9.7	440	10 Q7XXP2	Q7xxp2 avena sativ
38	228.5	9.6	451	10 Q64470	Q64470 arabidopsis
39	228	9.6	464	10 Q9MAP9	Q9map9 arabidopsis
40	227.5	9.6	433	10 Q9FI78	Q9fi78 arabidopsis
41	227.5	9.6	433	10 Q8LFT5	Q8lft5 arabidopsis
42	225	9.5	445	10 Q24645	Q24645 dianthus ca
43	222.5	9.4	441	10 Q9FLM5	Q9flm5 arabidopsis
44	222.5	9.4	457	10 Q94CD1	Q94cd1 arabidopsis
45	222	9.4	572	10 Q64549	Q64549 arabidopsis

ALIGNMENTS

RESULT 1

Q9FVF1 ID Q9FVF1 PRELIMINARY; PRT; 452 AA.  
AC Q9FVF1;  
DT 01-MAR-2001 (TREMREL. 16, Created)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)  
DE Alcohol acyltransferase.  
GN AAT.  
OS Fragaria ananassa (Strawberry).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;  
OC eurosids I; Rosales; Rosaceae; Rosoideae; Fragaria.  
OX NCBI\_TaxID=3747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fruit;  
RX MEDLINE=2021765; PubMed=10810141;  
RA Aharoni A., Keizer L.C.P., Boumeester H.J., Sun Z.,  
RA Alvarez-Huerta M., Verhoeven H.A., Blaas J.,  
RA Van Houtwelingen A.M.L., De Vos R.C.H., Van der Voet H.,  
RA Jansen R.C., Guis M., Mol J., Davis R.W., Schena M., van Tunen A.J.,  
RA O'Connell A.P.O.;  
RT "Identification of the SAA7 gene involved in strawberry flavor  
biogenesis by use of DNA microarrays.";  
RL Plant Cell 12:647-662(2000).  
EMBL; AF193789; AAG13130.1; -;  
DR GO; GO:0008415; F:acyltransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR InterPro; IPR003480; Transferase.  
DR Pfam; PF02459; Transferase; 1.  
KW Acyltransferase; Transferase.  
SQ SEQUENCE 452 AA; 50701 MW; 3D14F3BA732B135E CRC64;

Query Match 100.0%; Score 2373; DB 10; Length 452;  
Best Local Similarity 100.0%; Pred. No. 8.6e-208;  
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEKIEVINSKTIKPTSTSTPLQPKYKTLIDQLTPPAYVPVFYPTIDHFNLPQTLA 60



QY 4 IEVSINSKHTIKPSTSTPLQ--YKLTLLDQTPPAYVPIVFFYPITDHDNLPQTAD 61  
 DB 3 LEITVTSQELVSPSPNHNPPCHHLSFLDQAPLPIFMPFFFY-----HN-----KTNLSD 55  
 QY 62 -----LQALSETLTYPLSLGRVKN--LYIDDFFEGVPYLEARVNCMDTDFLRLRK 112  
 DB 56 KERSDHKISLSLSEILNLYPLAGRIKNSGDVVVNCNDV--GVSVKAKACNNSQILENPN 113  
 QY 113 IECLNEFVPIKPSMAISDERYPPLGVQNVDP--SGIAIGVSVSHKLDGGTADCFELKS 171  
 DB 114 ---PRLNKLHPFERFEVSD--VPLT-VQLTFFECGLGALGLSHKLDALSGLLIFVNS 167  
 QY 172 WGAVERGCRNIHPSLSAALLPPRDLPEKYVDOME--ALWFAKGVKATRRFVFGVKA 230  
 DB 168 WAAFAFGQTDILITSF-DLAKWFFPCD-----IENLNWAGITENIVTRFVLRSS 220  
 QY 231 ISSIOEAKSESVPKPSRVHVTGFLWKHLIAASRALTSGLTSTRLSIAQAQVNLTRMN 290  
 DB 221 VESLRETFSGNKIRATRVLEVLVSFVWSRFMAST---NHDDKTGKIYTLIHPVNLBRQAD 277  
 QY 291 METVLDNATGNLFWAAQAILELSHTTPEI-----SOLKCLDYNLLNGSVKQCGNDYFETP 346  
 DB 278 PD-IPDNFEGN-----IMRFSTVTPMMIINENDEEKASLDVQMBEIRKIDAVVYKXL 329  
 QY 347 KCKEGYGRMCEYLDFTWSSMEPADDIYLFSSWTNFFNPL---DFGWRGTSWIGVAGKI 403  
 DB 330 Q-EDNRG-----HLEFLNKQASGFVNGEIVSFSFSLCKEPVYVYADFGWGPVNVASA---381  
 QY 404 ESASCKFIILVPTQCGSGIEAWNLEEKWAMLEQDPHFLALASPKTLI 452  
 DB 382 RMSYKNLVAFIDTKEGDGIEAMINLDQDMSRFEADBEILLRYVSSNPSPVMV 432  
 RESULT 11  
 Q9FLW4 ID Q9FLW4 PRELIMINARY; PRT; 428 AA.  
 AC Q9FLW4; RC STRAIN=Columbia;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Acetyl-CoA:benzylalcohol acetyltransferase-like protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98290546; PubMed=9628582;  
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RA "Structural analysis of Arabidopsis thaliana chromosome 5. IV.  
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen  
 RT physically assigned pl and TAC clones.";  
 RL DNA Res. 5:41-54 (1998).  
 DR EMBL; AB009056; BAB08720.1; -  
 DR InterPro; IPR003480; Transferase.  
 DR Pfam; PF02458; Transferase; 1.  
 SQ SEQUENCE 428 AA; 47788 MW; 6CFF3F61F5703C4C CRC64;  
 Query Match 19.2%; Score 454.5; DB 10; Length 428;  
 Best Local Similarity 28.7%; Pred. No. 1.2e-32;  
 Matches 133; Conservative 87; Mismatches 192; Indels 51; Gaps 17;  
 QY 4 IEVSINSKHTIKPSTSTPLQYKLTLLDQTPPAYVPIVFFYPITDHDNLPQTADLR 63  
 DB 1 MKLELLSKEVIXPA-SPNHLQTLISLFDQFLSTVSAIFFY---DDHSNQEDIQRLX 55  
 QY 64 QALSETLTYPLSLGRVKNLYIDDFFEGVPYLEARVNCMDTDFLR-LRKIECLNEFVPI 122  
 DB 56 SSLSQTLSLFLYPLAGQIKGVTWCHNCDEGALFTEARAEIFLSDFLRNPSDADLIQKFI-- 113

QY 123 KPFSMAISDERYPPLGVQNVDP--SGIAIGVSVSHKLDGGTADCFELKSWGAVRGCRE 181  
 DB 114 ---VSPHAPETWPLLVKVIKPKDGFVAVSVSHKICDAASLSFVCSWTSKASKYAD 171  
 QY 182 NIHPSLSAALFPPRD---DLPEKYVDOMEALWFAKGVKATRRFVFGVKAISIODEA 238  
 DB 172 T-VDPEF-VGADFPADPDSIEPPLLVHE-----TKSKTKRFVFGSLMIEKLNRA 221  
 QY 239 KS-ESVPSKPSRVHVTGFLWKHLIAASRALTSGLTSTRLSIAQAQVNLTRMNMETVLDN 297  
 DB 222 SSKRRVQATRIESTITALLRCMTKAGHSKSGVKGEFAIT---QTMOLRPRVSSSLSPHK 278  
 QY 298 ATCKLFWAAQAILELSHTTPEI SOL--KLCCLDYNLLNGSVKQCGNDYFETFKGKEG----351  
 DB 279 AIGN-FFPLPKESSESKEWEIEETVSKLQKQTELINRDNSEDASKSVEAKERIASA 337  
 QY 352 -YGRMCEYLDFTWSSMEPADDIYLFSSWTNFFNPLDFGWRGTSWIGVAGKIESASCK 409  
 DB 338 MLSSICE-----ISPMEYAVSSWCRMSFYEAENFGWGPVNVAP-----DSVDKT 383  
 QY 410 FILLVPTQCGSGIEAWNLEEKWAMLEQDPHFLALASPKTLI 452  
 DB 384 QVVIMDSKDSGVEARVTLPTDVMKYEHDSLELVTAIPSESI 426  
 RESULT 12  
 Q23943 ID Q23943 PRELIMINARY; PRT; 99 AA.  
 AC Q23943; RC STRAIN=cv. Reine des Vallées;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Ripening-induced protein (Fragment).  
 OS Fragaria vesca (Woodland strawberry).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Rosales; Rosaceae; Rosoideae; Fragaria.  
 OX NCBI\_TaxID=57918;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Reine des Vallées;  
 RA Nam Y.W., Ticht L., Leperlier M., Querq B., Marty I., Lelievre J.M.;  
 RA "Isolation and characterization of cDNAs from genes differentially  
 RT expressed during ripening of wild strawberry (Fragaria vesca L.).";  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ001450; CAA04771.1; -  
 DR InterPro; IPR003480; Transferase;  
 DR Pfam; PF02458; Transferase; 1.  
 FT NON TER 1  
 SQ SEQUENCE 99 AA; 11336 MW; 3B567A4617095858 CRC64;  
 Query Match 19.0%; Score 450; DB 10; Length 99;  
 Best Local Similarity 87.5%; Pred. No. 3.3e-33;  
 Matches 84; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 QY 357 EYLDFTWSSMEPADDIYLFSSWTNFFNPLDFGWRGTSWIGVAGKIESASCKFIILVPT 416  
 DB 1 EYLDFTWSSMEPADDIYLFSSWTNFFNPLDFGWRGTSWIGVAGKIESASCKFIILVPT 60  
 QY 417 QCGSGIEAWNLEEKWAMLEQDPHFLALASPKTLI 452  
 DB 61 PCDTGIEAWNLEEKWAMLEQDPHFLALASPKTLI 96  
 RESULT 13  
 Q23393 ID Q23393 PRELIMINARY; PRT; 435 AA.  
 AC Q23393; RC STRAIN=cv. Reine des Vallées;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein (HSR201 like protein).  
 OX AT4G15400.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:24 ; Search time 42.3213 Seconds  
(without alignments)  
2176.455 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKVLVCGVCHSD.....ERVVKDVRFRFVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_29Jan04:\*
- 1: Geneseq1980s:\*
  - 2: Geneseq1990s:\*
  - 3: Geneseq2000s:\*
  - 4: Geneseq2001s:\*
  - 5: Geneseq2002s:\*
  - 6: Geneseq2003as:\*
  - 7: Geneseq2003bs:\*
  - 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1758	100.0	326	3 AAB36445	Strawberr
2	1758	100.0	326	3 AAY79662	Strawberr
3	1007	57.3	355	6 ADA38387	Soybean c
4	991	56.4	333	3 AAB36444	Strawberr
5	991	56.4	333	3 AAY79661	Strawberr
6	979	55.7	362	5 AAU80013	Sinapyl a
7	979	55.7	362	5 AAU79807	Sinapyl a
8	979	55.7	362	7 ADD33896	Quaking a
9	960	54.6	357	3 AAG29470	Arabidops
10	960	54.6	357	5 ABB33285	Herbicida
11	959	54.6	359	5 ABB33286	Herbicida
12	881	50.1	360	3 AAG40179	Arabidops
13	881	50.1	360	5 ABB33303	Herbicida
14	876	49.8	360	3 AAG25506	Arabidops
15	871	49.5	306	3 AAG29471	Arabidops
16	869	49.4	360	6 ADA38379	Soybean c
17	832	47.3	371	6 ADA38389	Wheat cin
18	827	47.0	283	3 AAB36447	Strawberr
19	827	47.0	283	3 AAY79664	Strawberr
20	824.5	46.9	278	3 AAB36446	Strawberr
21	824.5	46.9	278	3 AAY79663	Strawberr
22	819.5	46.6	375	3 AAG32129	Arabidops
23	819.5	46.6	375	5 ABB91832	Herbicida
24	802	45.6	376	3 AAG32132	Arabidops
25	800	45.5	361	6 ADA38383	Corn cinn

26	795	45.2	349	3 AAG36699	Arabidops
27	794	45.2	370	5 AAU75089	Ryegrass
28	790	44.9	366	6 ADA38385	Rice cinn
29	759.5	43.2	312	5 AAU75090	Ryegrass
30	732.5	41.7	364	6 ADA38377	Rice cinn
31	727	41.4	263	3 AAG29472	Arabidops
32	725	41.2	370	2 AAY05666	Maize cin
33	717	40.8	358	6 ADA38381	Wheat cin
34	716	40.7	358	2 AAY05668	Maize cin
35	709	40.3	359	2 AAY05667	Maize cin
36	690.5	39.3	358	2 AAR57415	Cinnamyl
37	679	38.6	357	5 AAU80016	Coniferyl
38	679	38.6	357	7 ADD93903	Quaking a
39	665	37.8	360	6 ADA38407	Wheat cin
40	661	37.6	407	5 AAU75088	Ryegrass
41	657	37.4	361	7 ADC68470	Lolium pe
42	652	37.3	361	7 ADC68387	S. arundi
43	652	37.1	361	7 ADC68386	Lolium pe
44	648	36.9	358	6 ADA38405	Soybean c
45	646	36.7	354	3 AAG38003	Arabidops

ALIGNMENTS

RESULT 1

AAB36445

ID AAB36445 standard; protein; 326 AA.

XX AC AAB36445;

XX AC

DT 28-FEB-2001 (first entry)

DE Strawberry alcohol dehydrogenase protein SEQ ID NO:7B.

KW Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase;  
KW aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;  
KW pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;  
KW alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;  
KW food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;  
KW confectionery; flavouring; oral medication; vitamin; aroma; beverage;  
KW alcohol; scent; fragrance; perfume; cosmetic; suspension aid;  
KW aluminium salt; anti-perspirant; pharmaceutical; cleaning product;  
KW insect pheromone; dye carrier; solvent; insect repellent; miticide;  
KW scabicide; plasticiser; deodorant.

XX OS Fragaria x ananassa.

XX OS WC2000032789-A1.

XX PD 08-JUN-2000.

XX PF 02-DEC-1999; 99WO-NL000737.

XX PR 02-DEC-1998; 98EF-00204018.

XX PR 12-MAR-1999; 99EF-00200739.

XX PA (CPRO-) CPRO-DLO CENT PLANTENVERDELINGS REPROD.

XX PI Aharoni A, Luecker J, Vertoeven HA, Van Tunen AJ, O'Connell AP;

XX PI WPI; 2000-412335/35.

XX PI N-PSDB; AAC64772.

XX PT A new DNA sequence encoding a polypeptide with alcohol acyl transferase  
XX activity for producing and regulating aromatic and/or aliphatic ester  
XX formation in microorganisms, plant cells or plants.

XX PS Claim 32; Page 89-90; 163pp; English.

XX CC The present invention describes nucleotide sequences with thiolase,  
XX alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
XX aminotransferase and esterase activities, which are involved in the

CC biosynthetic pathway for aliphatic and/or aromatic ester production in  
 CC fruit. The nucleotide sequences can be inserted into the genome of a  
 CC fruit-producing plant to regulate aliphatic and/or aromatic ester  
 CC formation. Aromatic and/or aliphatic esters in microorganisms, plant  
 CC cells or plants are produced by inserting thiolase, alcohol acyl  
 CC transferase, alcohol dehydrogenase, pyruvate decarboxylase,  
 CC aminotransferase and esterase nucleotide sequences into the genome and  
 CC feeding the microorganism or plant with alcohol, aldehydes, alpha-keto  
 CC acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and  
 CC their proteins can be used in the processed food industry as food  
 CC additives to enhance the flavour of syrups, ice-creams, frozen desserts,  
 CC yoghurts and confectionery. They are used as flavouring agents for oral  
 CC medications and vitamins; provide flavour and aroma in beverages,  
 CC including alcohol; enhance or reduce fruit flavour, aroma, fragrance or  
 CC scent; enhance the flavour or aroma of natural, synthetic or artificial  
 CC products; for the production of novel combinations of artificial flavour  
 CC substances; as antibacterial or anti-fungal agents; as fragrance or  
 CC perfumes in cosmetics, creams, sun-protectant products, hair  
 CC conditioners, lengthening agents and fixatives in perfumes, suspension  
 CC aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning  
 CC products, personal care products and animal care products; as  
 CC disinfectant additives; as degreasing solvents for electronics; as insect  
 CC pheromones; and as dye carriers, solvents, insect repellents, miticides,  
 CC scabicides, plasticisers and deodorants. The present sequence represents  
 CC the specifically claimed strawberry alcohol dehydrogenase

XX Sequence 326 AA;

Query Match 100.0%; Score 1758; DB 3; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 4e-171;  
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGCCKVKKFSWRQ 60  
 DB 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGCCKVKKFSWRQ 60

QY 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHFIVRIP 120  
 DB 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHFIVRIP 120

QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWEPFRFRPPLNLPGLWGSRLQS 180  
 DB 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWEPFRFRPPLNLPGLWGSRLQS 180

QY 181 LVPLIKEGSGYGTSPALMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240  
 DB 181 LVPLIKEGSGYGTSPALMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240

QY 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETOEMIDFAAENHITADIEVIPID 300  
 DB 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETOEMIDFAAENHITADIEVIPID 300

QY 301 YLNTAMERVVKVQVRFVIDVENTL 326  
 DB 301 YLNTAMERVVKVQVRFVIDVENTL 326

RESULT 2

AA79662  
 ID AA79662 standard; protein; 326 AA.

XX AA79662;

AC 12-SEP-2003 (revised)  
 DT 29-AUG-2000 (first entry)

XX Strawberry alcohol dehydrogenase SLF193 (C-terminal sequence).

XX Strawberry; alcohol dehydrogenase; fruit; ripening; ester; flavour;  
 XX aroma; transgenic plant.

OS Fragaria x ananassa.

XX

PN BP1006190-A1.

XX 07-JUN-2000.

XX 02-DEC-1998; 98EP-00204018.

XX 02-DEC-1998; 98EP-00204018.

XX (CPRO-) CPRO-DLO CENT PLANTENVREDELINGS REPROD.

XX Verhoeven HA, Van Tunen AJ, Aharoni A, Luecker J, O'Connell AP;

XX MPI; 2000-378264/33.

XX N-PSDB; AAA27672.

XX New polynucleotides encoding enzymes from the biosynthetic pathway for  
 PT aromatic and/or aliphatic ester production in fruit used to modify plant  
 PT flavors.

XX Claim 36; Page 81-82; 116pp; English.

XX The present sequence is that of the C-terminal region of strawberry cv.  
 CC Elsanta alcohol dehydrogenase SLF193, an enzyme that shows upregulated  
 CC expression in ripening fruit. The invention relates to DNA sequences (see  
 CC AAA27666-78) encoding enzymes (see AAY29656-68) involved in the metabolic  
 CC pathway leading to the formation of aliphatic and/or aromatic esters in  
 CC ripening fruit. The enzymes have alcohol acyl transferase, alcohol  
 CC dehydrogenase, pyruvate decarboxylase, thiolase or aminotransferase  
 CC activity. Expression vectors comprising the DNA sequences may be used to  
 CC regulate ester formation in fruit. Genetically modified plants, plant  
 CC cells and microorganisms can be used to produce esters. The DNA  
 CC sequences, polypeptides and antibodies are also used to screen fruit: for  
 CC volatile ester compounds; for quality such as flavour, fragrance, aroma,  
 CC scent, texture or shape; to distinguish between cultivars and varieties;  
 CC and to monitor harvest time, post-harvest quality, shelf-life, timing of  
 CC pesticide application, and resistance capacity based on volatile ester  
 CC profiles. (Updated on 12-SEP-2003 to standardise CS field)

XX Sequence 326 AA;

Query Match 100.0%; Score 1758; DB 3; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 4e-171;  
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGCCKVKKFSWRQ 60

DB 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGCCKVKKFSWRQ 60

QY 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHFIVRIP 120

DB 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHFIVRIP 120

QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWEPFRFRPPLNLPGLWGSRLQS 180

DB 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWEPFRFRPPLNLPGLWGSRLQS 180

QY 181 LVPLIKEGSGYGTSPALMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240

DB 181 LVPLIKEGSGYGTSPALMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240

QY 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETOEMIDFAAENHITADIEVIPID 300

DB 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETOEMIDFAAENHITADIEVIPID 300

QY 301 YLNTAMERVVKVQVRFVIDVENTL 326

DB 301 YLNTAMERVVKVQVRFVIDVENTL 326

RESULT 3

ADA38387

ID ADA38387 standard; protein; 355 AA.

XX



CC substances; as antibacterial or anti-fungal agents; as fragrance or  
 CC perfumes in cosmetics, creams, sun-protectant products, hair  
 CC conditioners, lengthening agents and fixatives in perfumes, suspension  
 CC aids for aluminum salts in anti-perspirant pharmaceuticals, cleaning  
 CC products, personal care products and animal care products; as  
 CC disinfectant additives; as degreasing solvents for electronics; as insect  
 CC pheromones; and as dye carriers, solvents, insect repellents, miticides,  
 CC scabicides, plasticizers and deodorants. The present sequence represents  
 CC the specifically claimed strawberry alcohol dehydrogenase  
 XX  
 XX

Sequence 333 AA;

Query Match 56.4%; Score 991; DB 3; Length 333;  
 Best Local Similarity 64.4%; Pred. No. 1.7e-92;  
 Matches 204; Conservative 25; Mismatches 56; Indels 32; Gaps 7;  
 QY 1 ETGATDVRFKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGCVKKFK-SWR 59  
 Db 34 ETGEKDVTFKVMYCGICHSDLHMWKEWGFSTYPLVPGHEIVGEVTEVGSNVQKFKVGR 93  
 QY 60 QQRCLWHLGRRLPTCNCIHLENYCPNLIQTYGKYDGTMTYGGYNNMVTDEHFIVRI 119  
 Db 94 VGVGIVGSCR-SCNCTDHLNYPCKQLITYGAKYDGTITTYGSDIMVADEHFIVRI 152  
 QY 120 PNPLDGAAPLLCAGITTYSPWRYGGLDKPMHLGVEMPRFRSRPPLNPLGLWSRLQ 179  
 Db 153 PNPLDGAAPLLCAGITTYSPRYGGLDKPMHLGVEMPRFRSRPPLNPLGLWSRLQ 196  
 QY 180 SLVPLLIKEGG-----SYGTSP-----ALMH-----SLRTDODOMEAAMSTMDGIIDTVP 225  
 Db 197 HVAVKFAKAMGVKVTIVTSPKKEBEARKHLGADSLVSRDODQQAALGTMDGIIDTWS 256  
 QY 226 AVRPLEPLISLLKTKGVVTVGIAVQPLDLPVFPPLIGKMKVAGSAIGMKETQEMIDPA 285  
 Db 257 AQHPLPLIGLLKSHGKLVWVGAPEKPLELPVFPPLMGRKMWAGSGIGMKETQEMIDPA 316  
 QY 286 AEHNITADIEVIPIDYL 302  
 Db 317 AKHNITADIEVIPIDYL 333

RESULT 5

AAU79661

ID AAU79661 standard; protein; 333 AA.

XX

AC AAU79661;

XX

DT 12-SEP-2003 (revised)

DT 29-AUG-2000 (first entry)

XX

DE Strawberry alcohol dehydrogenase SLB39.

XX

XX Strawberry; alcohol dehydrogenase; fruit; ripening; ester; flavour;

KW aroma; transgenic plant.

XX

OS Fragaria x ananassa.

XX

XX EPI006190-A1.

XX

XX 07-JUN-2000.

XX

PF 02-DEC-1998; 98EP-00204018.

XX

XX 02-DEC-1998; 98EP-00204018.

XX

XX (CPRO-) CPRO-DLO CENT PLANTENVERDEDELINGS REPROD.

XX

XX Verhoeven HA, Van Tunen AJ, Aharoni A, Luecker J, O'Connell AP;

XX

XX WPI; 2000-378264/33.

XX

XX N-PSDB; AAA27671.

XX

PT New polynucleotides encoding enzymes from the biosynthetic pathway for

PT aromatic and/or aliphatic ester production in fruit used to modify plant  
 PT flavors.  
 XX

Claim 35; Page 79-80; 116pp; English.

XX The present sequence is that of alcohol dehydrogenase SLB39 of strawberry  
 CC cv. Elsanta, an enzyme involved in the metabolic pathway leading to the  
 CC production of volatile ester compounds in fruit, and in the ripening  
 CC process. The invention relates to DNA sequences (see AAA27666-78)  
 CC encoding enzymes (see AAY29656-68) involved in the metabolic pathway  
 CC leading to the formation of aliphatic and/or aromatic esters in ripening  
 CC fruit. The enzymes have alcohol acyl transferase, alcohol dehydrogenase,  
 CC pyruvate decarboxylase, thiolase or aminotransferase activity. Expression  
 CC vectors comprising the DNA sequences may be used to regulate ester  
 CC formation in fruit. Genetically modified plants, plant cells and  
 CC microorganisms can be used to produce esters. The DNA sequences,  
 CC polypeptides and antibodies are also used to screen fruit: for volatile  
 CC ester compounds; for quality such as flavour, fragrance, aroma, scent,  
 CC texture or shape; to distinguish between cultivars and varieties; and to  
 CC monitor harvest time, post-harvest quality, shelf-life, timing of  
 CC pesticide application, and resistance capacity based on volatile ester  
 CC profiles. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 333 AA;

Query Match 56.4%; Score 991; DB 3; Length 333;  
 Best Local Similarity 64.4%; Pred. No. 1.7e-92;  
 Matches 204; Conservative 25; Mismatches 56; Indels 32; Gaps 7;  
 QY 1 ETGATDVRFKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGCVKKFK-SWR 59  
 Db 34 ETGEKDVTFKVMYCGICHSDLHMWKEWGFSTYPLVPGHEIVGEVTEVGSNVQKFKVGR 93  
 QY 60 QQRCLWHLGRRLPTCNCIHLENYCPNLIQTYGKYDGTMTYGGYNNMVTDEHFIVRI 119  
 Db 94 VGVGIVGSCR-SCNCTDHLNYPCKQLITYGAKYDGTITTYGSDIMVADEHFIVRI 152  
 QY 120 PNPLDGAAPLLCAGITTYSPWRYGGLDKPMHLGVEMPRFRSRPPLNPLGLWSRLQ 179  
 Db 153 PNPLDGAAPLLCAGITTYSPRYGGLDKPMHLGVEMPRFRSRPPLNPLGLWSRLQ 196  
 QY 180 SLVPLLIKEGG-----SYGTSP-----ALMH-----SLRTDODOMEAAMSTMDGIIDTVP 225  
 Db 197 HVAVKFAKAMGVKVTIVTSPKKEBEARKHLGADSLVSRDODQQAALGTMDGIIDTWS 256  
 QY 226 AVRPLEPLISLLKTKGVVTVGIAVQPLDLPVFPPLIGKMKVAGSAIGMKETQEMIDPA 285  
 Db 257 AQHPLPLIGLLKSHGKLVWVGAPEKPLELPVFPPLMGRKMWAGSGIGMKETQEMIDPA 316  
 QY 286 AEHNITADIEVIPIDYL 302  
 Db 317 AKHNITADIEVIPIDYL 333

RESULT 6

AAU80013

ID AAU80013 standard; protein; 362 AA.

XX

AC AAU80013;

XX

DT 15-JUL-2002 (first entry)

XX

DE Sinapyl alcohol dehydrogenase (SAD).

XX

XX Plant; aspen; phenylpropanoid pathway; agronomic; lignin; paper;

KW 4-coumarate-CoA ligase; 4CL; conferyl aldehyde 5-hydroxylase; CAL5H;

XX S-adenosyl-L-methionine-dependent; SAM; Ald6MT; transgenic; grass;

XX 5-hydroxyconiferaldehyde O-methyltransferase; cellulose; pulp;

KW conferyl alcohol dehydrogenase; CAD; sinapyl alcohol dehydrogenase; SAD;

XX syringyl; guaiacyl; agriculture.

XX

OS Populus tremuloides.

XX



Db 35 TGEDVRFKVLKYGICHSDLHSIKNDWGFMSYPLVPGHEIVGEVTEVGSVKVNVGDKV 94  
 QY 61 -----GRCWLHGLRPTCENCIHLENYCPNLIQTYGSKYVGTMTYGGYNNMVTDEH 114  
 Db 95 GWGCLVGAC-----HSCSCANDLENYCPKMLTYASIHDTYGGYSDHVMANER 147  
 QY 115 FIVRIPDNPLDGAAPLLCAGITTYSPWRYGDKPGMLGVSWPRFRSRPPLNPLGLW 174  
 Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGDLDFGKHIGI-----VGLG 193  
 QY 175 GSRLOSLVPLLIKEGGS-----YGTSPALMSHLRT-----DODQMEAMSTWDGI 220  
 Db 194 G--LGHVAVKFAKAFGSKVTIVSTSPSKKEALKNFAGDSFLVSRDQEQMGAAGTLDGI 251  
 QY 221 IDTVPAVRPLEPLISLILKNGKVVTVGIAVQPLDLFVPELIIGRKMVAGSAIGMKETQE 280  
 Db 252 IDTVSAVHPLEPLFGLLKHGKLLILVGAPEKPLELPAFSLIAGRKIVAGSIGMKETQE 311  
 QY 281 MIDFAAEHNITADIEVPIIDYLTAMERVVKKDVRFRFVIDVENTL 326  
 Db 312 MIDFAAEHNITADIEVISTDYLTATIERLAKNDVRYRFVIDVENTL 357

## RESULT 8

ADD93896  
 ID ADD93896 standard; protein; 362 AA.

AC ADD93896;

DT 29-JAN-2004 (first entry)

DE Quaking aspen sinapyl alcohol dehydrogenase (SAD).

QW Quaking aspen; Plant; enzyme; 4-coumarate-CoA ligase; 4CL;

KW conferyl aldehyde 5-hydroxylase; CAL5H;

KW SAM-dependent 5-hydroxyconiferaldehyde O-methyltransferase;

KW S-adenosyl-L-Methionine; ALD5MT; conferyl alcohol dehydrogenase; CAD;

KW sinapyl alcohol dehydrogenase; SAD; transgenic; agronomic property;

KW lignin; cellulose; syringyl/guaiacyl lignin ratio; growth; wood quality;

KW stress resistance; sterility; grain yield; nutritional value; S/G ratio;

KW paper delignification; pulp manufacture; grass digestibility.

XX Populus tremuloides.

OS US2002138870-A1.

PN 26-SEP-2002.

PD 06-MAR-2002; 2002US-00091009.

PF 05-SEP-2000; 2000US-0230086P.

PR 05-SEP-2001; 2001US-00947027.

XX (UNMT ) UNIV MICHIGAN TECHNOLOGICAL.

PA Chiang VLC, Li L;

PI WPI; 2003-843045/78.

XX N-PSDB; ADD93895.

XX Genetic transformation of plants, useful for altering lignin and

PT cellulose contents, by introducing genes from the phenylpropanoid

PT pathways, also new transgenic plants.

XX Disclosure; SEQ ID NO 2; 55pp; English.

XX The invention relates to the genetic transformation of a plant

CC simultaneously with several genes from the phenylpropanoid pathways by

CC incorporating into the genome the genes for 4-coumarate-CoA ligase (4CL),

CC conferyl aldehyde 5-hydroxylase (CAL5H), S-adenosyl-L-Met-dependent 5-

CC hydroxyconiferaldehyde O-methyltransferase (ALD5MT), conferyl alcohol

CC dehydrogenase (CAD) and sinapyl alcohol dehydrogenase (SAD), their

CC fragments or combination to produce plants with altered agronomic traits.  
 CC Also included are preparing plant cells that contain, in the genome,  
 CC several DNA constructs containing one or more of the specified genes,  
 CC preparing transgenic plants with altered lignin or cellulose contents (by  
 CC regenerating plants from the cells and expressing the construct in cells  
 CC of the regenerated plant), transgenic plants produced by method above and  
 CC their progeny, plants that include in the genome a construct containing  
 CC at least one of the specified gene (linked to a promoter and a  
 CC terminator) a set of DNA constructs (each containing a promoter,  
 CC terminator and at least one of the specified genes or their fragments)  
 CC and the set of above or a single construct containing the 4CL gene, where  
 CC incorporated into a plant genome. The method is used to transform plants,  
 CC particularly trees but also forage crops and monocotyledons, to alter  
 CC their agronomic properties, especially lignin and cellulose contents,  
 CC syringyl/guaiacyl (S/G) lignin ratio, growth, wood quality, stress  
 CC resistance, sterility, grain yield and nutritional value, particularly to  
 CC increase S/G ratio (this simplifies delignification in paper and pulp  
 CC manufacture) and to increase digestibility of grasses. The present  
 CC sequence represents quaking aspen SAD.

XX Sequence 362 AA;

Query Match 55.7%; Score 979; DB 7; Length 362;

Best Local Similarity 56.6%; Pred. No. 3.2e-91;

Matches 196; Conservative 39; Mismatches 67; Indels 44; Gaps 6;

QY 2 TGATDVRFKVLKYGICHSDLHSIKNDWGFMSYPLVPGHEIVGEVTEVGSVKVNVGDKV 60

Db 35 TGEDVRFKVLKYGICHSDLHSIKNDWGFMSYPLVPGHEIVGEVTEVGSVKVNVGDKV 94

QY 61 -----GRCWLHGLRPTCENCIHLENYCPNLIQTYGSKYVGTMTYGGYNNMVTDEH 114

Db 95 GWGCLVGAC-----HSCSCANDLENYCPKMLTYASIHDTYGGYSDHVMANER 147

QY 115 FIVRIPDNPLDGAAPLLCAGITTYSPWRYGDKPGMLGVSWPRFRSRPPLNPLGLW 174

Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGDLDFGKHIGI-----VGLG 193

QY 175 GSRLOSLVPLLIKEGGS-----YGTSPALMSHLRT-----DODQMEAMSTWDGI 220

Db 194 G--LGHVAVKFAKAFGSKVTIVSTSPSKKEALKNFAGDSFLVSRDQEQMGAAGTLDGI 251

QY 221 IDTVPAVRPLEPLISLILKNGKVVTVGIAVQPLDLFVPELIIGRKMVAGSAIGMKETQE 280

Db 252 IDTVSAVHPLEPLFGLLKHGKLLILVGAPEKPLELPAFSLIAGRKIVAGSIGMKETQE 311

QY 281 MIDFAAEHNITADIEVPIIDYLTAMERVVKKDVRFRFVIDVENTL 326

Db 312 MIDFAAEHNITADIEVISTDYLTATIERLAKNDVRYRFVIDVENTL 357

## RESULT 9

AAG29470

ID AAG29470 standard; protein; 357 AA.

AC AAG29470;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 35071.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136382P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142380P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147453P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148550P.  
PR 13-AUG-1999; 99US-0148634P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149456P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151045P.  
PR 27-AUG-1999; 99US-0151086P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-015138P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.



```
DT 31-MAY-2002 (first entry)
XX AC
DE Herbicidally active polypeptide SEQ ID NO 2497.
XX KW
XX Herbicidal; plant; agriculture; herbicide.
XX OS
XX Arabidopsis thaliana.
XX FN
XX WO200210210-A2.
XX PD
XX 07-FEB-2002.
XX XX
XX 28-AUG-2001; 2001WO-EP009892.
XX PF
XX 28-AUG-2001; 2001WO-EP009892.
XX XX
XX (FARB ) BAYER AG.
XX PA
XX Tietjen K, Weidner M;
XX PI
XX WPT; 2002-269010/31.
XX DR
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms.
XX XX
XX Claim 5; SEQ ID NO 2497; 26lpp + Sequence Listing; English.
XX PS
XX
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
XX for herbicidally active compounds, comprising aligning and comparing
XX nucleic acid or amino acid sequences from plant with nucleic acid or
XX amino acid sequences from non-plant organisms using suitable search
XX parameters, where plant sequences having an E-value greater by a factor
XX of 3 than the E-value of most similar non-plant sequences are selected.
XX The polypeptides or nucleic acids encoding them are useful for
XX identifying modulators. The identified modulators are useful as
XX herbicides
XX CC
XX Sequence 359 AA;
XX SQ
XX
Query Match 54.6%; Score 959; DB 5; Length 359;
Best Local Similarity 55.1%; Pred No. 3.6e-89;
Matches 193; Conservative 39; Mismatches 68; Indels 50; Gaps 5;
QY 1 ETGADVRKFLVGVCHSDHMAKNDWGTSTYPIVPGHVLGVVTVGVCKVKFKSWRQ 60
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
30 ETGEKDVRFKVLFCGICHSDLHMVKNWGMSTYPLVPGHVLGVVTVGVCAKVKFKTGEK 89
QY 61 -----GRCLHGRLEPTCNCIHLENYCPNLIQTYGKYDGTWYGYSNMWTDE 113
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
90 VGVGLVSSC-----GSCDSCTEGMEYCPKSIQTYGFFYDNTITYGYSIDHVCVEE 142
QY 114 HFTVRIPDNLPLDGAAPLFCAGITTYSPWRYGLDKPGMHLGVWSPRRFRSRPPLNPLGL 173
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
143 GFVIRIPDNLPLDGAAPLFCAGITTYSPWRYGLDKPGMHLGVWSPRRFRSRPPLNPLGL 188
QY 174 WGRSLQSLVPLIKEGSGSYGTSPLMHS-----LRTDQDQMEAAAMST 216
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
189 GG-----LGHVGVKFAKAMGKTVTVISTSEKRDENRLGADAFVSRDPKQIKDAMGT 243
QY 217 MDGIIDTVPAVRPLEPLISLTKTKGVTVGVIAVQPIQLDPVFPPLIIGKRWAGSAGSMK 276
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
244 MDGIIDTVSATHSLLPLLLGLLKHGKGLVWGAPEKPLEPVPMPPLIFFRKRWYMGSMIGSIK 303
QY 277 ETQMIDFAAEHNTTADIEVPIIDVLTAMERVVKQVRPFVIDVENTIL 326
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
304 ETQMIDWAGKNTTADIELISADYVNTMERLEKADVRYRFVIDVANTIL 353
XX
RESULT 12
AAG40179
ID AAG40179 standard; protein; 360 AA.
```

```
XX AC
XX AAG40179;
XX DT
XX 18-OCT-2000 (first entry)
XX DE
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49819.
XX KW
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS
XX Arabidopsis thaliana.
XX FN
XX EP1033405-A2.
XX PD
XX 06-SEP-2000.
XX XX
XX 25-FEB-2000; 2000EP-00301439.
XX PF
XX 25-FEB-1999; 99US-0121825P.
XX PR
XX 05-MAR-1999; 99US-0123180P.
XX PR
XX 09-MAR-1999; 99US-0123548P.
XX PR
XX 23-MAR-1999; 99US-0125788P.
XX PR
XX 25-MAR-1999; 99US-0126264P.
XX PR
XX 29-MAR-1999; 99US-0126785P.
XX PR
XX 01-APR-1999; 99US-0127462P.
XX PR
XX 06-APR-1999; 99US-0128234P.
XX PR
XX 08-APR-1999; 99US-0128714P.
XX PR
XX 16-APR-1999; 99US-0129845P.
XX PR
XX 19-APR-1999; 99US-0130077P.
XX PR
XX 21-APR-1999; 99US-0130449P.
XX PR
XX 23-APR-1999; 99US-0130510P.
XX PR
XX 28-APR-1999; 99US-0130891P.
XX PR
XX 30-APR-1999; 99US-0131449P.
XX PR
XX 30-APR-1999; 99US-0132048P.
XX PR
XX 04-MAY-1999; 99US-0132407P.
XX PR
XX 04-MAY-1999; 99US-0132484P.
XX PR
XX 05-MAY-1999; 99US-0132485P.
XX PR
XX 06-MAY-1999; 99US-0132486P.
XX PR
XX 06-MAY-1999; 99US-0132487P.
XX PR
XX 07-MAY-1999; 99US-0132863P.
XX PR
XX 11-MAY-1999; 99US-0134256P.
XX PR
XX 14-MAY-1999; 99US-0134218P.
XX PR
XX 14-MAY-1999; 99US-0134219P.
XX PR
XX 14-MAY-1999; 99US-0134221P.
XX PR
XX 18-MAY-1999; 99US-0134370P.
XX PR
XX 19-MAY-1999; 99US-0134941P.
XX PR
XX 20-MAY-1999; 99US-0135124P.
XX PR
XX 21-MAY-1999; 99US-0135353P.
XX PR
XX 24-MAY-1999; 99US-0135629P.
XX PR
XX 25-MAY-1999; 99US-0136021P.
XX PR
XX 27-MAY-1999; 99US-0136392P.
XX PR
XX 28-MAY-1999; 99US-0136782P.
XX PR
XX 01-JUN-1999; 99US-0137222P.
XX PR
XX 03-JUN-1999; 99US-0137528P.
XX PR
XX 04-JUN-1999; 99US-0137502P.
XX PR
XX 07-JUN-1999; 99US-0137724P.
XX PR
XX 08-JUN-1999; 99US-0138094P.
XX PR
XX 10-JUN-1999; 99US-0138540P.
XX PR
XX 10-JUN-1999; 99US-0138847P.
XX PR
XX 14-JUN-1999; 99US-0139119P.
XX PR
XX 16-JUN-1999; 99US-0139452P.
XX PR
XX 16-JUN-1999; 99US-0139453P.
XX PR
XX 17-JUN-1999; 99US-0139432P.
XX PR
XX 18-JUN-1999; 99US-0139454P.
XX PR
XX 18-JUN-1999; 99US-0139455P.
XX PR
XX 18-JUN-1999; 99US-0139456P.
XX PR
XX 18-JUN-1999; 99US-0139457P.
XX PR
XX 18-JUN-1999; 99US-0139458P.
XX PR
XX 18-JUN-1999; 99US-0139459P.
XX PR
XX 18-JUN-1999; 99US-0139460P.
XX PR
XX 18-JUN-1999; 99US-0139461P.
```

Query Match 50.1%; Score 881; DB 3; Length 360;

```
QY 217 MDGIITDTPAVRPLEPLISLLKTKNGKVVTVGIAVOPDLDPTEPLIIGRWVAGSAIGMK 276
Db 248 MDYIIDTISAVHALYPLGLKLVNGKLTALGPEKPLPELPMFPLVIGRWVAGSDVGGMK 307
QY 277 ETQEMIDFAAENHNTADIEVIPIDYINTAMERVKVKKDVRFRFVIDVENTL 326
Db 308 ETQEMLDFCAKHNITADIELIKMDEINTAMERLAKSDVRYRVIDVANSL 357

RESULT 13
ABB93303
AC ABB93303;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2514.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
FN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP009892.
XX
PR 28-AUG-2001; 2001WO-EP009892.
XX
PA (FARB ) BAYER AG.
XX
PI Tietjen K, Weidier M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
comprising aligning and comparing nucleic acid or amino acid sequences
from plant with nucleic acid or amino acid sequences from non-plant
organisms.
XX
PS Claim 5; SEQ ID NO 2514; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016;
for herbicidally active compounds, comprising aligning and comparing
nucleic acid or amino acid sequences from plant with nucleic acid or
amino acid sequences from non-plant organisms using suitable search
parameters, where plant sequences having an E-value greater by a factor
of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
identifying modulators. The identified modulators are useful as
herbicides
XX
SQ Sequence 360 AA;

Query Match 50.1%; Score 881; DB 5; Length 360;
Best Local Similarity 49.1%; Pred. No. 3.6e-81;
Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGANDVRFKVLVCGVCHSDHMAKNDWTSYTPVPGHELVGVTVENGCKVKFKFSWRQ 60
Db 34 DNGENDVTVKILFCVGVCHDTHATKNDMGYSYVPVPGHEIVGIAIKVGNKNTKFKEGDR 93
QY 61 -----GRCWLHGLRLPTCNCIHLHNYCNPLIQTYSKYIDGTMTYGGYSNNMVTDE 113
Db 94 VGVGVISGSC-----QSCSCDQDLNFCYQMSFTYNAIGSDGKTKYGGYSENIIVDQ 146
QY 114 HFIVRIPNPLDGNAPLLCAGITITYSPWRYGLDKPGMHGLGVWFRFRSRPPLNLPGL 173
Db 147 RFLRFPENPLSDSGAPLLCAGITITYSPMKYYGWTAGKHLGV-----AGL 192
```

```
QY 174 WGSRLQSLQVPPLIKGGSGYGTSPALMHS-----LLRTDQDQWEAMST 216
Db 193 GG-----LGHVAVKLGKAFGLKVTWISSSTKAEAINHLGADSELTWTDPOKMGAAIGT 247
QY 217 MDGIITDTPAVRPLEPLISLLKTKNGKVVTVGIAVOPDLDPTEPLIIGRWVAGSAIGMK 276
Db 248 MDYIIDTISAVHALYPLGLKLVNGKLTALGPEKPLPELPMFPLVIGRWVAGSDVGGMK 307
QY 277 ETQEMIDFAAENHNTADIEVIPIDYINTAMERVKVKKDVRFRFVIDVENTL 326
Db 308 ETQEMLDFCAKHNITADIELIKMDEINTAMERLAKSDVRYRVIDVANSL 357

RESULT 14
AAG25506
ID AAG25506 standard; protein; 360 AA.
XX
AC AAG25506;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 29600.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128645P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134258P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
```

PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142190P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148313P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 16-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151530P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 28-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160982P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 49.8%; Score 876; DB 3; Length 360;  
Best Local Similarity 48.9%; Pred. No. 1.2e-80;  
Matches 171; Conservative 52; Mismatches 77; Indels 50; Gaps 5;



PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 03-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 49.5%; Score 871; DB 3; Length 306;  
Best Local Similarity 54.7%; Pred. No. 3e-80;  
Matches 176; Conservative 43; Mismatches 65; Indels 38; Gaps 5;

QY 23 MAKNDWGTSTYPIVPGHEHVGWVTEVGCKVKPKSW-ROGRCLHGRRLRPTCNCIHHLE 81  
DB 1 MAKNEWGLTITPLVPGHEHVGWVTEVGAKVKFNAGDKVGVGMAGSCR-SCDSCNDGDE 59  
QY 82 NYCPLNIQYGSKYDGTWYTGYSNNMVTDEHFIIVRIIPDKLELDGAAPLLCAGITTYSP 141  
DB 60 NYCPRMILTSGAKNFDDTWHGYSYSDHVMVCAEDFIIRIPDNLELDGAAPLLCAGITTYSP 119  
QY 142 WRYYGLDKQMLGVWEPFRFRSRPPNLPGLWGSRLQSLVPLIXEGGSGYGTSPALMHS 201  
DB 120 MKYHGLDKPQMHIGV-----VGLGG-----LGHVAVKFAKAMGTKVTVIST 160  
QY 202 -----LLRTDDOMEAAMSTMDGIIDTVPAVRPLEPLISILKTKNGKV 244  
DB 161 SERKRDEAVTRLGADAFVSRDEKQMKDAMGTMDGIIDIVSATHPLPLILGLLKNKGKLV 220  
QY 245 TVGIAVOPLDLVPFPLIIGRKXVAGSAIGMKETQEMIDFAAEHNTTADIEVIPIDYANT 304  
DB 221 MVGAPAEPLPLVPFPLIFGRKXVAGSAIGMKETQEMIDFAAEHNTTADIEVIPIDYANT 280  
QY 305 AMERVVKQVRFVRFVVDVENTL 326  
DB 281 AMERLAKADVRFVRFVVDVANTM 302

Search completed: July 6, 2004, 13:36:45  
Job time : 44.3213 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:34:55 ; Search time 13.8278 Seconds  
(without alignments)  
1217.121 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758  
Sequence: 1 ETGATDVRFKVLVCGVCHSD.....ERVVKDVRFRFVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1007	57.3	355	4	US-09-501-115-12
2	869	49.4	360	4	US-09-501-115-4
3	832	47.3	371	4	US-09-501-115-14
4	800	45.5	361	4	US-09-501-115-8
5	790	44.9	366	4	US-09-501-115-10
6	732.5	41.7	364	4	US-09-501-115-2
7	717	40.8	358	4	US-09-501-115-6
8	665	37.8	360	4	US-09-501-115-32
9	648	36.9	358	4	US-09-501-115-30
10	630.5	35.9	353	4	US-09-328-352-7473
11	579.5	33.0	410	4	US-09-252-991A-25647
12	578	32.9	357	4	US-09-501-115-20
13	567.5	32.3	358	4	US-09-489-039A-10508
14	559	31.8	292	4	US-09-501-115-28
15	379	21.6	360	4	US-09-501-115-34
16	366.5	20.8	374	4	US-09-433-039A-8925
17	356	20.3	358	4	US-09-328-352-4988
18	316	18.0	344	4	US-09-134-001C-5343
19	310	17.6	349	2	US-07-857-224B-95
20	303	17.2	186	4	US-09-501-115-16
21	299.5	17.0	150	4	US-09-501-115-24
22	298.5	17.0	172	4	US-09-501-115-46
23	297.5	16.9	465	4	US-09-252-991A-18919
24	296	16.8	155	4	US-09-501-115-48
25	273.5	15.6	375	4	US-09-134-000C-5937
26	270	15.4	206	4	US-09-540-236-3768
27	265	15.1	346	4	US-09-489-039A-9885

28 261 14.8 344 2 US-07-857-224B-93 Sequence 93, Appl  
29 255 14.5 344 2 US-07-857-224B-92 Sequence 92, Appl  
30 248 14.1 128 4 US-09-501-115-44 Sequence 44, Appl  
31 245.5 14.0 152 4 US-09-501-115-40 Sequence 40, Appl  
32 242 13.8 346 2 US-07-857-224B-96 Sequence 96, Appl  
33 240 13.7 97 4 US-09-501-115-42 Sequence 42, Appl  
34 238 13.5 125 4 US-09-501-115-36 Sequence 36, Appl  
35 238 13.5 343 3 US-09-363-189B-4 Sequence 4, Appl  
36 232 13.2 345 2 US-07-857-224B-94 Sequence 94, Appl  
37 216 12.3 116 4 US-09-540-236-3774 Sequence 3774, Ap  
38 209 11.9 344 4 US-08-311-731A-92 Sequence 82, Appl  
39 203.5 11.6 352 4 US-09-504-358-18 Sequence 18, Appl  
40 203.5 11.1 103 4 US-09-504-314-18 Sequence 18, Appl  
41 195.5 11.1 103 4 US-09-501-115-22 Sequence 22, Appl  
42 192.5 10.9 352 4 US-09-648-004-24 Sequence 24, Appl  
43 190 10.8 114 4 US-09-501-115-38 Sequence 38, Appl  
44 181 10.3 111 4 US-09-501-115-18 Sequence 18, Appl  
45 173 9.8 152 4 US-09-501-115-26 Sequence 26, Appl

#### ALIGNMENTS

RESULT 1  
US-09-501-115-12  
; Sequence 12, Application US/095031115  
; Patent No. 6552249  
; GENERAL INFORMATION:  
; APPLICANT: Caboon, Rebecca E.  
; APPLICANT: Fager, Gary M.  
; APPLICANT: Rafalski, Antoni  
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs  
; FILE REFERENCE: BBL328 US NA  
; CURRENT APPLICATION NUMBER: US/09/501.115  
; EARLIER FILING DATE: 2000-02-09  
; EARLIER APPLICATION NUMBER: 60/119,585  
; EARLIER FILING DATE: 1993-February-10  
; NUMBER OF SEQ ID NOS: 48  
; SEQ ID NO 12  
; SOFTWARE: Microsoft Office 97  
; LENGTH: 355  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-501-115-12

Query Match 57.3%; Score 1007; DB 4; Length 355;  
Best Local Similarity 59.6%; Pred. No. 1.2e-100;  
Matches 201; Conservative 39; Mismatches 73; Indels 24; Gaps 5;

Qy 1 ETGATDVRFKVLVCGVCHSDIHMANKNDWGTSTYPIVPGHELGVGVTVGCKVKXPK-SNR 59  
Db 28 ETSEKDLVFKVQYQYCGICHSDIHMANKNDWGTSTYPIVPGHELGVGVTVGCKVKXPKVQK 87  
Qy 60 QGRCLMGRRLRPTCNCIHLNVCNLIQYGYKYDGTMTYGYSSNNMVTDEHFVRI 119  
Db 88 VGVCMIGSCR-SCSCDENLNYCPKMLTYGVKFDGTHGYSIDMVADEHFVRI 146  
Qy 120 PDNLPLDGAAPLLCAGITTYSPWYYGLDKFGEHMLGVWEPFRSPRPPLNPLGLWGSRLQ 179  
Db 147 PDNLPLDGAAPLLCAGITTYSPWYYGLDKFGEHMLGVWEPFRSPRPPLNPLGLWGSRLQ 194  
Qy 180 SLVPPLIKEGSGVTSPLMHSLRT-----DQDQEAAMSTWDGIIIDTVPAVRP 229  
Db 195 KFAKALGANVTIVISTSPNKKKEATENTGADSFVVSREQDQQAQVMTDGIIDTVSAVHP 254  
Qy 230 LEPLISLTKNGKVVTVYGVIAVQVLDLVPFLLITGRKMWAGSAIGMKETOEMIDFAEHN 289  
Db 255 LVPLIGLKHGKLVVMVGAPEKPELFPVFLLMGRKMWGSSIGMKETOEMIDFAKHG 314  
Qy 290 ITADIEVPIIDYLNLTAMERVVKDVRFRFVIDVENTL 326  
Db 315 VKPDIEVPIIDYLNLTALERLAKADVRFVIDIGNTL 351

```
RESULT 2
US-09-501-115-4
; Sequence 4, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BB1328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Glycine max
US-09-501-115-4

Query Match      49.4%; Score 869; DB 4; Length 360;
Best Local Similarity 51.3%; Pred. No. 1.1e-85;
Matches 174; Conservative 47; Mismatches 88; Indels 30; Gaps 4;

QY 1 ETGATDVRPKVLYCGVCHSDIHMAKNDWGSTYPIVPGHVLGVVTEVCGCKVKKSWRQ 60
DB 34 ENGVDVDTLKILFCGVCHSDHLTKNDWGSTFTYVPGHVLGVVTKVGNVKNKRVGDK 93
QY 61 GRCLMHGLRPTCNCIHLHENYCPNLQTYGSKYDGTMTYGGYSNNMTDEHFIYRI 120
DB 94 VGVGVIVSEKCECQDLESYCPREPTYNSPYDGTGRTKRGYSNIMVHQRYLRRP 153
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 180
DB 154 ENLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 197
QY 181 LVPPPLKEGGSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDTVA 226
DB 198 VAIKLAKAFGLKLLSSSPENQAEADRLGADSLVSSDPAKMKVAGTMDYIIDTISA 257
QY 227 VRPLPLSLKTKNGKVVTVGIAVQPLDLPVFPPLIIGRMVAGSAGKVKKFSW-RQ 286
DB 258 VHSPLPLGLLKLNGKLVTVGLPNKPLEPIPLVAGRKLIGSNFGGIKETQEMLDFA 317
QY 287 EHNITADIEVIPIDYINTAMERVVKQVFRFRVIDVENT 325
DB 318 KHNITADIELKMDQINTAMERLSKADVKYRFVIDVANS 356

RESULT 3
US-09-501-115-14
; Sequence 14, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BB1328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-501-115-14

Query Match      47.3%; Score 832; DB 4; Length 371;
Best Local Similarity 48.8%; Pred. No. 1.2e-81;
Matches 168; Conservative 52; Mismatches 84; Indels 40; Gaps 6;

QY 2 TGATDVRPKVLYCGVCHSDIHMAKNDWGSTYPIVPGHVLGVVTEVCGCKVKKFSW-RQ 60
DB 40 TGDDDDVIXIKLYCGVCHSDLSIKNDWKNKAKYPMIPGHEIAGEVTEVGNVTKFKAGDRV 99
QY 61 G-RCMLHGLRPTCNCIHLHENYCPNLQTYGSKYDGTMTYGGYSNNMTDEHFIYRI 119
DB 100 GVGCMVNS--CQSCSCDKGPFENHCPGMFTTNSVDRDGTGRTGTHGGYSNNVWVHERFWVF 157
QY 120 PDLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 179
DB 158 PDAMPDLKGAFLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 201
QY 180 SLVPPPLKEGGSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIID 222
DB 202 --HVAVKFGKAFGMKVTVISSSPGKQKQEALERLGADAFVVSADAKATMTMDGIIN 258
QY 223 TVPVRPLPLSLKTKNGKVVTVGIAVQPLDLPVFPPLIIGRMVAGSAGKVKKFSW 282
DB 259 TVSANVPMAFLPGLLKNPKMIMVGLPEKPIEVPPFALVARNKTLGSCIGMRDQEWML 318
QY 283 DFAAEHNITADIEVIPIDYINTAMERVVKQVFRFRVIDVENTL 326
DB 319 DLAAKHGVTADIEVGAENVNTAMERLAKADVYRFVIDIANL 362

RESULT 4
US-09-501-115-8
; Sequence 8, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BB1328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (39)
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (78)
US-09-501-115-8

Query Match      45.5%; Score 800; DB 4; Length 361;
Best Local Similarity 47.7%; Pred. No. 3.5e-78;
Matches 165; Conservative 52; Mismatches 83; Indels 46; Gaps 7;

QY 2 TGATDVRPKVLYCGVCHSDIHMAKNDWGSTYPIVPGHVLGVVTEVCGCKVKKFSW-RQ 60
DB 33 TGDDDDVIXIKLYCGVCHSDLSIKNDWKNKAKYPMIPGHEIAGEVTEVGNVTKFKAGDRV 92
QY 61 G-RCMLHGLRPTCNCIHLHENYCPNLQTYGSKYDGTMTYGGYSNNMTDEHFIYRI 119
DB 93 GVGCMVNS--CQSCSCDEDFPNSCRAVIFTTNSVDRDGTGTYGGYSNNVWVHERFWVF 150
QY 120 PDLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGL 173
DB 151 PDAMPDLKGAFLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGL 199
```



```

QY 59 RQRCWLHGRRLPPTCENCITHLENYCPNLQIOTYGSKYVD---GTMTYGGYSNNMVTDEH 114
DB 93 FVDS-----RSCDSCKSYESCQPLVETNSVSLDDGCGATTGGSFSDALVHVOR 145
QY 115 FIVRIPDNLPDGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPPLNPLGL- 173
DB 146 YVVRVPASLPAPGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPPLNPLGL- 194
QY 174 -----WGRSLQSLVPLPKKGGSGYTSPLMHSLRLTDQDOMEAAMSTMDGLIDTV 224
DB 195 HLAVERFGKAPGMKVTVISTSLGRDEALGRGADAP-LVSRDPEQRAAAGTLGVIDTV 253
QY 225 PAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLIIGRKMVAGSAIGMKETQRMIDF 284
DB 254 SADHPVPLDLLKPMQGVVVGLPTKLOVPAFSLVAGGKRVAGSAGSGVCEQAMLDLF 313
QY 285 AAENHNTADIEVIPIDYLNMTAMERVVKVDVFRFVIDVENT 325
DB 314 AGEHGITADVEVGVMDYNTATQRLERNDVRYFRFVVDVASS 354

```

## RESULT 8

```

US-09-501-115-32
; Sequence 32, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: B31328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; EARLIER FILING DATE: 2000-02-09
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-501-115-32

```

```

Query Match 37.8%; Score 665; DB 4; Length 360;
Best Local Similarity 40.6%; Pred. No. 1.6e-63;
Matches 141; Conservative 57; Mismatches 97; Indels 52; Gaps 7;

QY 1 ETGATDVRPKVLYCGVCHSDIHMAKNDGTSTPIVPGHVLGVTVGVCKVKKFSWR- 59
DB 32 KTGPDVVLYKVCIGCHTDHVQKNDLGASKYPMVPGHEVGVGVGVGVGVGVGVGV 91
QY 60 -----QGRCLHGRRLPPTCENCITHLENYCPNLQIOTYGSKYVDGTMTYGGYSNNMVTDE 113
DB 92 VGVGVVGC-----RDCRCKANVQYCNKLTWSNDVTDKPTGGFASAMVVDQ 144
QY 114 HFIVRIPDNLPDGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPPLNPLGL 173
DB 145 KPVVKIPAGLAPEQAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPPLNPLGL 193
QY 174 WGRSLQSLVPLPKKGGSGYTSPLMHSLRLTDQDOMEAAMSTMDGLIDTV 216
DB 194 GHMG-----VYKAKSMGHVTVTSSNKKRAEMDGLGADAYLVSSDTPDQMAAADS 245
QY 217 MDGIDTVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLIIGRKMVAGSAIGM 275
DB 246 LDYIITUPAKHLEPYLALKQDGLKVLGVIAPLSF-VSPMWLGRKKTITGSGFISGM 304
QY 276 KETQRMIDFAAENHNTADIEVIPIDYLNMTAMERVVKVDVFRFVIDV 322
DB 305 DETEVLQFCVDKGLTSQIEVVKMDYVQAFERLERNDVRYFRFVVDV 351

```

## RESULT 9

```

Query Match 35.9%; Score 630.5; DB 4; Length 353;
Best Local Similarity 39.9%; Pred. No. 8.6e-60;

```

```

US-09-501-115-30
; Sequence 30, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: B31328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; EARLIER FILING DATE: 2000-02-09
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 30
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (201)
US-09-501-115-30

```

```

Query Match 36.9%; Score 648; DB 4; Length 358;
Best Local Similarity 41.0%; Pred. No. 1.1e-61;
Matches 141; Conservative 58; Mismatches 103; Indels 42; Gaps 9;

QY 2 TGATDVRPKVLYCGVCHSDIHMAKNDGTSTPIVPGHVLGVTVGVCKVKKFSWRQ 61
DB 32 TGPDDVYIKVHYCGICHSDLHQIKNDLGMSNYPMPVPGHEVGVGVGVGVGVGVGV 88
QY 62 RCWLHGRRLPPTCENCITHLENYCPNLQIOTYGSKYVDGTMTYGGYSNNMVTDEHFI 118
DB 89 ELVGVGLLVGCKNQCPQDIENTCYCKKINSNDVYVDGKTFQGFATMTVEQKRFVVK 148
QY 119 IPDNLPDGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPPLNPLGL- 176
DB 149 IPEGLAPQVAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPPLNPLGL- 197
QY 177 RQSLVPLPKKGGSGYTSPLMHSLRLTDQDOMEAAMSTMDGLIDTV 222
DB 198 KXAXAL-----GHVTVTSSDKKKQEALEHLGADQYLVSSDATAMQEAADSLDIID 250
QY 223 TVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLIIGRKMVAGSAIGMKETQ 281
DB 251 TVPUGHLEPYLALKQDGLKVLGVIAPLSF-VSPMWLGRKKTITGSGFISGMKETE 309
QY 282 IDFAAENHNTADIEVIPIDYLNMTAMERVVKVDVFRFVIDVENT 325
DB 310 LEFMKEKGLSSMIEMVNDYINKAFERLEKNDVRYFRFVVDVVKGS 353

```

## RESULT 10

```

US-09-328-352-7473
; Sequence 7473, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7473
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7473

```



Query Match	31.8%;	Score	559;	DB	4;	Length	292;
Best Local Similarity	41.6%;	Pred.	No. 3.7e-52;				
Matches	124;	Conservative	44;	Mismatches	88;	Indels	42;
Gaps	9;						

  

QY	1	ETCATDVRKVLVCGYCHSDIHMAKNDNGTSTPIVPGHGLVGVVTEGCKVKPKSW	R	59
Db	16	ENGNDWTIKLYCGICHDTDLHYAKNENGITPVVPGHEIIGVTKVGRDVKVGFKEGR	75	
QY	60	QG--RCLWGLRLRPTCENCIIHLENYCPNLTQITGSKSYDGTMTITGGYSNMWTDHFIVR	118	
Db	76	VGVGCLAAACLE--CBHCKTDQENYCEKLFQVYNGVFDGSIITYGYSQIFVADYRYVVH	133	
QY	119	IDPNLPLDGAAPLLCAGIITYSPWRYGL-DKPGMHLGVWEPFRFRSRPDLNPLGLNGSR	177	
Db	134	IDENLMDAAAPLLCAGITVPNPLKDHLSVASPGKIGV-----VGLGG--	177	

Search completed: July 6, 2004, 13:40:16  
Job time : 14.8278 secs

QY 1 ETGATDVREKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGC

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:38:51 ; Search time 36.036 Seconds

(without alignments)  
2816.033 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKVLYCGVCHSD.....ERVVKDVRFRVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/2/pubpaa/2CT_NEW_PUB.pap:*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pap:*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pap:*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pap:*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pap:*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pap:*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pap:*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pap:*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pap:*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pap:*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pap:*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pap:*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pap:*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pap:*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pap:*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pap:*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	57.3	359	12	US-10-424-599-266181
2	1007	57.3	355	14	US-10-357-886-12
3	1007	57.3	362	12	US-10-424-599-261443
4	979	55.7	362	9	US-09-947-150-2
5	979	55.7	362	9	US-09-947-027-2
6	979	55.7	362	13	US-10-091-009-2
7	958	54.5	361	12	US-10-424-599-266973
8	929.5	52.9	360	12	US-10-425-114-55801
9	884	50.3	356	12	US-10-424-599-152603
10	884	50.3	365	12	US-10-425-114-36617
11	884	50.3	365	12	US-10-425-114-54709
12	880	50.1	360	12	US-10-424-599-152604
13	880	50.1	366	12	US-10-425-114-44445
14	869	49.4	360	14	US-10-357-886-4
15	842	47.9	364	12	US-10-424-599-205468

```
16 842 47.9 381 12 US-10-425-114-51033 Sequence 51033, A
17 832 47.3 371 14 US-10-357-886-14 Sequence 14, Appl
18 810 46.1 381 12 US-10-425-114-64190 Sequence 64190, A
19 800 45.5 361 14 US-10-357-886-8 Sequence 8, Appl
20 790 44.9 366 14 US-10-357-886-10 Sequence 10, Appl
21 779 44.3 428 15 US-10-437-963-118472 Sequence 118472,
22 755.5 43.0 360 16 US-10-437-963-204767 Sequence 204767,
23 741 42.0 378 12 US-10-425-114-39429 Sequence 39429, A
24 732.5 41.7 364 14 US-10-357-886-2 Sequence 2, Appl
25 725 41.2 370 14 US-10-361-460-10 Sequence 10, Appl
26 725 41.2 406 12 US-10-425-114-56149 Sequence 56149, A
27 724.5 41.2 373 12 US-10-425-114-40577 Sequence 40577, A
28 718.5 40.9 370 12 US-10-425-114-45843 Sequence 45843, A
29 718.5 40.9 370 12 US-10-425-114-62384 Sequence 62384, A
30 717 40.8 358 14 US-10-357-886-6 Sequence 6, Appl
31 716 40.7 358 14 US-10-361-460-12 Sequence 12, Appl
32 716 40.7 378 12 US-10-425-114-59052 Sequence 59052, A
33 716 40.7 380 12 US-10-425-114-39395 Sequence 39395, A
34 716 40.7 390 12 US-10-425-114-43324 Sequence 43324, A
35 716 40.7 391 12 US-10-425-114-39519 Sequence 39519, A
36 714.5 40.6 379 12 US-10-425-114-71567 Sequence 71567, A
37 713.5 40.6 336 16 US-10-437-963-172457 Sequence 172457,
38 712.5 40.5 391 16 US-10-437-963-123886 Sequence 123886,
39 709 40.3 359 14 US-10-361-460-11 Sequence 11, Appl
40 707 40.2 410 16 US-10-437-963-176237 Sequence 176237,
41 697.5 39.7 329 12 US-10-425-114-54812 Sequence 54812, A
42 696 39.6 384 12 US-10-424-599-230622 Sequence 230622,
43 679 38.6 357 9 US-09-947-027-9 Sequence 9, Appl
44 679 38.6 357 13 US-10-091-009-9 Sequence 9, Appl
45 670.5 38.1 344 15 US-10-369-493-13927 Sequence 13927, A
```

#### ALIGNMENTS

#### RESULT 1

```
US-10-424-599-266181
; Sequence 266181, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266181
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_82383C.1.pap
US-10-424-599-266181
```

```
Query Match 57.3%; Score 1008; DB 12; Length 359;
Best Local Similarity 59.5%; Pred. No. 6.3e-97;
Matches 203; Conservative 42; Mismatches 64; Indels 32; Gaps 7;

Qy 1 ETGATDVRFKVLYCGVCHSDIHAKNDWGTSTSTPIVPGHVLGVVTVGCVKXKPK-SWR 59
34 EPKEKDVAFVLYCGI CHSDIHSIKNEWGTSTIYPMVPGHEVAGVTVGSKVPEKFGDK 93
60 QGRCLHGRSLPTCCNCIHLNLCNLYCPNLIOYGYGKYDGTMTYGYGYNMAYTDSHF-VRI 119
94 VGVCLVDSRC-TCQNCNDNLNLENTTYYGAKYRDGITYGYSDSYVADEHFVRI 152
120 PDNLPLDGAAPLLCAGITTYSPWRYGYLDKPKGMELGVWEPFRFRPPLNLPGLWGSRLQ 179
153 PDRLPLDGAAPLLCAGITTYSPWRYGYLDKPKGLVGV-----VGLGG--LG 196
```



Db 35 TGEEDVRFKVLYCGICHSDLSHDKNDWGFSPYPLVPGHGVTEVGVSKVKNVGVGDKV 94  
Qy 61 -----GRCWLHGRRLPTCENCIIHLENYCPMLIQTYGKYDGTMTYGGYSNNMTDEH 114  
Db 95 GVGCLVAGC-----HSCESCANDLENYCPKMLITYASIHGDTITYGGYSDDHVMANER 147  
Qy 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYGGLDKPGMHLGVWEPFRFRPLNLPGLW 174  
Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKYLFGLDPEPKHIGI-----VGLG 193  
Qy 175 GSRLQSLVPLIKEGGS-----YGTSPALMHSLLRT-----DQDQMEAMSTMDGI 220  
Db 194 G--LGHVAVKFAKFGSKVTYSTSPSKKEALKNFAGDSFLVSRDQEQMRAAGTLDDGI 251  
Qy 221 IDTVPAVRPLEPLISLLKNGKVVTVGIAVQPLDLVPFPLIIGRKWVAGSAIGMKETQE 280  
Db 252 IDTVSAVHPLPLFGLLKGSHGLILVGAPEKPLELPFAFSLIAGRKIVAGSGIGMKETQE 311  
Qy 281 MIDFAAENHITADIEVIPIDYLNATAMERVVKDVRFRFVIDVENTL 326  
Db 312 MIDFAAKHNITADIEVISTDYLNATIERLAKNDVRYRFRFVIDVGNL 357

## RESULT 5

US-09-947-027-2  
; Sequence 2, Application US/09947027  
; Patent No. US20020124281A1  
; GENERAL INFORMATION:  
; APPLICANT: Chiang, Vincent Lee C.  
; APPLICANT: Li, Laigeng  
; TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITION  
; FILE REFERENCE: 066040-9718  
; CURRENT APPLICATION NUMBER: US/09/947,027  
; PRIOR FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: 60/230,086  
; PRIOR FILING DATE: 2000-09-05  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 362  
; TYPE: PRT  
; ORGANISM: aspen populus tremuloides  
US-09-947-027-2

Query Match 55.7%; Score 979; DB 9; Length 362;  
Best Local Similarity 56.6%; Pred. No. 7.2e-94;  
Matches 196; Conservative 39; Mismatches 67; Indels 44; Gaps 6;

Qy 2 TGATDVRFKVLYCGVCHSDIHMKNWGTSTYPIVPGHVLGVWTEVGVCKVKFKSWRC- 60  
Db 35 TGEEDVRFKVLYCGICHSDLSHDKNDWGFSPYPLVPGHGVTEVGVSKVKNVGVGDKV 94  
Qy 61 -----GRCWLHGRRLPTCENCIIHLENYCPMLIQTYGKYDGTMTYGGYSNNMTDEH 114  
Db 95 GVGCLVAGC-----HSCESCANDLENYCPKMLITYASIHGDTITYGGYSDDHVMANER 147  
Qy 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYGGLDKPGMHLGVWEPFRFRPLNLPGLW 174  
Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKYLFGLDPEPKHIGI-----VGLG 193  
Qy 175 GSRLQSLVPLIKEGGS-----YGTSPALMHSLLRT-----DQDQMEAMSTMDGI 220  
Db 194 G--LGHVAVKFAKFGSKVTYSTSPSKKEALKNFAGDSFLVSRDQEQMRAAGTLDDGI 251  
Qy 221 IDTVPAVRPLEPLISLLKNGKVVTVGIAVQPLDLVPFPLIIGRKWVAGSAIGMKETQE 280  
Db 252 IDTVSAVHPLPLFGLLKGSHGLILVGAPEKPLELPFAFSLIAGRKIVAGSGIGMKETQE 311  
Qy 281 MIDFAAENHITADIEVIPIDYLNATAMERVVKDVRFRFVIDVENTL 326  
Db 312 MIDFAAKHNITADIEVISTDYLNATIERLAKNDVRYRFRFVIDVGNL 357

## RESULT 6

US-10-091-009-2  
; Sequence 2, Application US/10091009  
; Publication No. US20020138870A1  
; GENERAL INFORMATION:  
; APPLICANT: Chiang, Vincent Lee C.  
; APPLICANT: Li, Laigeng  
; TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND  
; TITLE OF INVENTION: COMPOSITION, AND  
; FILE REFERENCE: 066040-9718  
; CURRENT APPLICATION NUMBER: US/10/091,009  
; CURRENT FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: 09/947,027  
; PRIOR FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: 60/230,086  
; PRIOR FILING DATE: 2000-09-05  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 362  
; TYPE: PRT  
; ORGANISM: aspen populus tremuloides  
US-10-091-009-2

Query Match 55.7%; Score 979; DB 13; Length 362;  
Best Local Similarity 56.6%; Pred. No. 7.2e-94;  
Matches 196; Conservative 39; Mismatches 67; Indels 44; Gaps 6;

Qy 2 TGATDVRFKVLYCGVCHSDIHMKNWGTSTYPIVPGHVLGVWTEVGVCKVKFKSWRC- 60  
Db 35 TGEEDVRFKVLYCGICHSDLSHDKNDWGFSPYPLVPGHGVTEVGVSKVKNVGVGDKV 94  
Qy 61 -----GRCWLHGRRLPTCENCIIHLENYCPMLIQTYGKYDGTMTYGGYSNNMTDEH 114  
Db 95 GVGCLVAGC-----HSCESCANDLENYCPKMLITYASIHGDTITYGGYSDDHVMANER 147  
Qy 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYGGLDKPGMHLGVWEPFRFRPLNLPGLW 174  
Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKYLFGLDPEPKHIGI-----VGLG 193  
Qy 175 GSRLQSLVPLIKEGGS-----YGTSPALMHSLLRT-----DQDQMEAMSTMDGI 220  
Db 194 G--LGHVAVKFAKFGSKVTYSTSPSKKEALKNFAGDSFLVSRDQEQMRAAGTLDDGI 251  
Qy 221 IDTVPAVRPLEPLISLLKNGKVVTVGIAVQPLDLVPFPLIIGRKWVAGSAIGMKETQE 280  
Db 252 IDTVSAVHPLPLFGLLKGSHGLILVGAPEKPLELPFAFSLIAGRKIVAGSGIGMKETQE 311  
Qy 281 MIDFAAENHITADIEVIPIDYLNATAMERVVKDVRFRFVIDVENTL 326  
Db 312 MIDFAAKHNITADIEVISTDYLNATIERLAKNDVRYRFRFVIDVGNL 357

## RESULT 7

US-10-424-599-266973  
; Sequence 266973, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic, David K  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 266973  
; LENGTH: 361  
; TYPE: PRT

```

; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_83098C.1.pep
US-10-424-599-266973

Query Match      54.5%; Score 958; DB 12; Length 361;
Best Local Similarity 56.6%; Pred. No. 1.2e-91;
Matches 194; Conservative 43; Mismatches 70; Indels 36; Gaps 6;

Qy 1  BTGATDVRKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVTEVGCCKVKFKSWRQ 60
Db 34  KTGKDVTFKVLVCGVCHSDHLKKNWSDSIYPLVPGREIVGEVTEVGSKVDKFKVGDK 93
Qy 61  -----GRCMLHGRLRPTCNCIHHLNENYCNLIQTGYSKYDGTMTYGGYSNNWVTD 113
Db 94  VAAGCLVGC-----HSCQCNVNLENYCCQVIPTYGAKYVDGTIIVGFSDFWVAD 146
Qy 114  HFTVIRPNLDGAPLLCAGITTYSPWRYYGLDKPGHGLGVPRPRSRPPLNLPOL 173
Db 147  HFVNIPSAIPDLAAAPLLCAGITTYGRLRYFGLDKPGMHLGV-----VGLGGL 195
Qy 174  WGSRLQSLVPPLLKGGSGVTSP-----ALMH-----SLRTDQDMBAAMSTMDGIIDT 223
Db 196  -GHLAVKFAKALGLKVTYISTSPKKNEAIQHGLGADFFVSRDQDMQAMCTLDGIIDT 254
Qy 224  VPAVRPLEPLISLLKTKNGKVVTVGIAVQPLDLVPFPLIIGRKWVAGSAIGMKETQEMID 283
Db 255  VSAVHPLMPLIDLKSHGKLVAVGAPEKPELELLPPLILGRKSIAGSYIGGIKETQEMIN 314
Qy 284  FAAENITADIEVIPDYLNTAMERVVKDVRFRFVIDVENTL 326
Db 315  FAEENVRPEIEVPMYVNTAMERLQKADVKYRFVIDIGNTL 357

RESULT 8
US-10-425-114-55801
; Sequence 55801, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55801
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPIC033A03_FLI.pep
US-10-425-114-55801

Query Match      52.9%; Score 929.5; DB 12; Length 360;
Best Local Similarity 56.2%; Pred. No. 1.1e-88;
Matches 190; Conservative 40; Mismatches 78; Indels 35; Gaps 6;

Qy 1  BTGATDVRKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVTEVGCCKVKFKSWRQ 60
Db 42  KTGKDVTFKVLVCGVCHSDHLKKNWSDSIYPLVPGREIVGEVTEVGSKVDKFKVGDK 101
Qy 61  -----GRCMLHGRLRPTCNCIHHLNENYCNLIQTGYSKYDGTMTYGGYSNNWVTD 113
Db 102  VAAGCLVGC-----HSCQCNVNLENYCCQVIPTYGAKYVDGTIIVGFSDFWVAD 154
Qy 114  HFTVIRPNLDGAPLLCAGITTYSPWRYYGLDKPGHGLGVPRPRSRPPLNLPOL 173
Db 114  HFVNIPSAIPDLAAAPLLCAGITTYGRLRYFGLDKPGMHLGV-----VGLGGL 195

; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_108825C.1.pep
US-10-424-599-152603

Query Match      50.3%; Score 884; DB 12; Length 356;
Best Local Similarity 50.6%; Pred. No. 6.9e-84;
Matches 173; Conservative 49; Mismatches 84; Indels 36; Gaps 3;

Qy 1  BTGATDVRKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVTEVGCCKVKFKSWRQ 60
Db 30  ENGVDVDTLKILFCGCHSDHLTKANDWGTFTYVVPVPGHEIVGVTEVGNVKNKVGDK 89
Qy 61  GRCMLHGRLRPTCNCIHHLNENYCNLIQTGYSKYDGTMTYGGYSNNWVTDHEFIVRIP 120
Db 90  VGVGVIVESCCEKCCQDLENYCPRPVFTYNSPYDGTGRTGGYSNIVVVHQRVLRFP 149
Qy 121  DNLPDGAAPLLCAGITTYSPWRYYGLDKPGHGLGVPRPRSRPPLNLPGLWGRLOS 180
Db 150  ENLPDAGAPLLCAGITTYSPMKYYGTEPGKHLGV-----AGLGG----- 190
Qy 181  LVPPLIKGGSGYGTSPALMHS-----LLRTDQDMBAAMSTMDGIIDT 223
Db 191  LGHVAIKLAKAFGLKVTYVTSPPNNKQAEADRLGADFFLVSDPAKMKAAALGTMDYIIDT 250
Qy 224  VPAVSPCEPLISLLKTKNGKVVTVGIAVQPLDLVPFPLIIGRKWVAGSAIGMKETQEMID 283
Db 251  ISAVHSLPLGLLKLNGKLVTVGLPNKPLEDIPFLVAGRLKLGSGNPGGLKETQEMID 310
Qy 284  FAAENITADIEVIPDYLNTAMERVVKDVRFRFVIDVENT 325
Db 311  FCGKENITADIELKMDQINTAMERLSEADVKYRFVIDVASS 352

RESULT 10
US-10-425-114-36617
; Sequence 36617, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
```



[illegible]

```

RESULT 13
US-10-425-114-44445
; Sequence 44445, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44445
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700954121 FL1.pep
US-10-425-114-44445

```

```

Query Match      50.1%; Score 880; DB 12; Length 366;
Best local Similarity 50.6%; Pred. No. 1.9e-83;
Matches 173; Conservative 49; Mismatches 84; Indels 36; Gaps 3;

Qy      1  ETGATDVRFKVLCVCGCHSDIHAKNDWGTSTVPIVPGHELGVVYTVGCKYKFKSWRQ 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      40  ENGVDVDTLKILFCGCHSDLTLKNDWGTFTVPVPGHEIVGVVTVKGVNNVKNFKYGDK 99
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      61  GROWLHGRLRPTCENCJLHLENYCPNLIQTGYGSKYYDGTMTYGGYSNNMVTDEHFIVRIP 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      100  VGVGVIVESCKESCCODLESICPRFVFNYSFYDGTGRTKGYSNIMVWHQRYVLREP 159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      121  DNLPLGAAPLLCAGITTYTPWRYTGDKPGMHLGVWEPRRFRPRPLNIPGLWGSRLQS 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      160  ENLPLGAGAPLLCAGITTYSPMKYTYGMPGKHGV-----AGLGG----- 200
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      181  LVFPLLKEGSGYGTSPALMHS-----LLRTDODOMEAAMSTMDGHIIDT 223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      201  LGHVAIKLAQFGLKVTVISSSPNKQAEADIRLGADSFVSSDPAPKMKVALGTMDYIIDT 260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      224  VPAVRPLEPLISLLKTKGVVTVGIAVQPLDLVPFPLIIIGKGVWAGSAIGMKETQEMID 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      261  ISAVHSLIPLGLLKLKNGKLVTVGLPNKPLELPFPLVAGRKLGGSNFGGIKETQEMLD 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      284  FAEHNITADIEVIPIDYLTAMERVVKDVRFRFVIDVENT 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      321  FCAKNHITADIELIKMDIINTAMERUSKADWKYRFVIDVANS 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14  
US-10-357-886-4  
; Sequence 4, Application US/10357886  
; Publication No. US20030159170A1  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Rebecca E.  
; APPLICANT: Pader, Gary M.  
; APPLICANT: Rafalski, Antoni  
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs  
; FILE REFERENCE: BB1328 US NA  
; CURRENT APPLICATION NUMBER: US/10/357,886

```

: CURRENT FILING DATE: 2003-02-04
: PRIOR APPLICATION NUMBER: US/09/501.115
: PRIOR FILING DATE: 2000-02-09
: PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/119,585
: PRIOR FILING DATE: EARLIER FILING DATE: 1999-February-10
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 4
: LENGTH: 360
: TYPE: PRT
: ORGANISM: Glycine max
: US-10-357-886-4

      Query Match          49.4%; Score 869; DB 14; Length 360;
      Best Local Similarity 51.3%; Pred. No. 2.6e-82;
      Matches 174; Conservative 47; Mismatches 88; Indels 30; Gaps 4;

Qy      1  ETGATDVRFKVLYCGVCHSDIHMAKNDMCTSTYPIVPGHLYGVVTEGCKVKFKFSWQ 60
Db      34  ENGVDVDVTLKILFCGVCSDLHTLKNMGFTTYPPVVPFGHEIVGVVTKVGNVKNFKVGVGK 93

Qy      61  GRCWLHGRLRPTCENCIHLENYCPMLIOTYGSKYDGTMTYGGYSYNNMTDEHFIVRIP 120
Db      94  VGVGVTVESCKECSQQDLESYCPRPVPTYNSPYYDGTGTRKRGYSNIMVWHQRYVLRFP 153

Qy      121  DNLPLDGAFLPLCAGTITTSRPRYCYGLDQPGMHLGVENPRFRSRPPLKPLGLWGSRLQS 180
Db      154  ENLPLDAGAPLLCAGTITTSRPMKYIGTEPFGKHLGV-----AGLGG--LGH 197

Qy      181  LVPPPLKEGG----SYGTSFALMHS-----LLRTDQDQMAAMSTMGIIDTVA 226
Db      198  VAIKAKAGLKLSSSSSENQAEADRLGADSFVSSDPKMKVALGTMVDYIDTISA 257

Qy      227  VRPLEPLISLLKNGKVVTVGIAVQPLDLPVFPFLIIGRWVAGSAIGGMKETQEMIDFAA 286
Db      258  VHSLIPLLGLLKLNGKLVTVGLPNKPLELPPIPLVAGRKLIGGSNFGGIGKETQEMLDPCA 317

Qy      287  EHNITADIEVPIDYINTAMERVVKQVRFPRVIDVENT 325
Db      318  KKNITADIELIKQDQINTAMERLSRADVKYRFVIDVANS 356

```

```

RESULT 15
US-10-424-599-205468
; Sequence 205468, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 205468
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_27566C.1.pap
US-10-424-599-205468

```

```

Query Match      47.9%; Score 842; DB 12; Length 364;
Best Local Similarity 49.6%; Pred. No. 1.9e-79;
Matches 171; Conservative 50; Mismatches 80; Indels 44; Gaps 6;

QY 3 GATDVFREFVLYGVCHSDHMAKNDWGTSTYPIVPQHELVGWTEGVCKVKPK----- 56
38 GNDITLDVLYSGICTLDLHWNDNFRISYIPWPGHEIVGKTVKGRVTKTSVWGDIAG 97
Db

```

```

Qy 57 -SWQGRCLHGRRLPTCENCIHLENYCPNLQIOTYGSYYDGTMTYGGYSNNMVTDEHF 115
Db 98 VGGSVGSC-----GSCSDCSNGFYVYCPXILTYSAHYDGTITQGGYSNNIYVDQNF 150
Qy 116 IVRIPDNLPDGAAPLLCAGITTYSPMYYYGLDKPGMHLGVEMPRFRSRPPLNLPGLWG 175
Db 151 VWLIPKSLPLDGAAPLLCAGITTYSPMYYYGLDKPGMHLGV-----VGLGG 196
Qy 176 SRLQSLVPPLIKEGSY-----GTSPALMHS-----LRTDQDQMEAMSTMDGII 221
Db 197 --LGHVAVKFAKAFGMHVTIVTSFSKKEBEALEKIGADEFEVSLDQOQLODARGTMDAIL 254
Qy 222 DTVPVPRLEPLISLLKTNKGKVTWVGINVPQPLDLPVPLIIGRKWAGSAIGMKETOEM 281
Db 255 DTVSANHSIQPLIALLKTSGLIIVGGPPSPLEVLAMPPLLGRKMIAGSAGGGRBIOEM 314
Qy 282 IDFAAEHNITADIEVIFIDYLNATAMERVVKDVRFRFVIDVENTL 326
Db 315 MDPFAAKHNITADVEVIMPMDYVNTAFERLEKNDVKYRFVIDVANTI 359

```

Search completed: July 6, 2004, 13:48:42  
Job time : 37.036 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:59 ; Search time 8.38046 Seconds  
(without alignments)  
2025.529 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKVLYCGVCHSD.....ERVVKDVRFRVIDVENTIL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1092	62.1	359	1 MTD_FRAAN	Q9ZRF1; ADH1_BACST
2	971	55.2	361	1 MTD_MESCR	P42327 bacillus st
3	963	54.8	337	1 MTD_PETCR	P25377 saccharomyc
4	960	54.6	357	1 MTD1_ARATH	O31186 rhizobium m
5	939	54.6	359	1 MTD2_ARATH	P07754 emericella
6	946	53.8	365	1 MTD_APIGR	P49383 kluyveromyc
7	884	50.3	360	1 MTDH_ARATH	O94038 candida alb
8	870.5	49.5	359	1 MTD3_STYHU	P41747 aspergillus
9	860.5	48.9	363	1 MTD1_STYHU	P43067 candida alb
10	833.5	47.4	354	1 MTD1_PICAB	Q9p6c8 neurospora
11	740	42.1	357	1 CAD2_PICAB	P20369 kluyveromyc
12	740	42.1	357	1 CAD7_PICAB	Q9p4c2 kluyveromyc
13	739	42.0	357	1 CADH_PINTA	
14	723	41.1	357	1 CADH_PINRA	
15	691.5	39.3	360	1 CAD1_ARACO	
16	686	39.0	357	1 CAD9_TOBAC	
17	681	38.7	357	1 CAD4_TOBAC	
18	666	37.9	357	1 CADH_POPEE	
19	656	37.3	361	1 CADH_LOLEP	
20	654	37.2	356	1 CAD2_EUCGU	
21	653	37.1	356	1 CADH_EUCGL	
22	646	36.7	357	1 CAD2_ARATH	
23	641	36.5	358	1 CADH_MESDA	
24	629	35.8	354	1 CAD1_EUCGU	
25	628	35.7	367	1 CADH_MAIZE	
26	621.5	35.4	355	1 CADH_EUCBO	
27	621	35.3	365	1 CAD1_ARATH	
28	611	34.8	365	1 CADH_SACOF	
29	599.5	34.1	346	1 ADH_MYCTU	
30	555.5	31.6	349	1 YAHK_ECOLI	
31	379	21.6	360	1 YW97_YEAST	
32	365.5	20.8	339	1 YJGB_ECOLI	
33	356.5	20.3	339	1 ADH3_BACST	

RESULT 1  
MTD\_FRAAN STANDARD; PRT; 359 AA.  
AC Q9ZRF1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol  
dehydrogenase).  
GN CAD.  
OS Fragaria ananassa (Strawberry).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids I; Rosales; Rosaceae; Rosoideae; Fragaria.  
OX NCBI\_TaxID=3747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Chandler;  
RT Medina-Escobar N., Caballero J.L., Munoz-Blanco J.;  
RT "Cloning, sequencing and temporal and spatial expression pattern of  
the cinnamyl alcohol dehydrogenase";  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step  
CC by which translocated mannitol is committed to central metabolism  
CC and, by regulating mannitol pool size, is important in regulating  
CC salt tolerance at the cellular level (By similarity).  
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.  
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
CC family.  
CC -!- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol  
CC dehydrogenase.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U63534; RAD10327.1; --  
DR InterPro; IPR002328; ADH\_zinc.  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR Pfam; PF00107; ADH\_zinc\_N; --  
DR PROSITE; PS00059; ADH\_ZINC; 1.  
KW Oxidoreductase; Zinc; Metal-binding; NAD.  
FT METAL 48 48 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 70 70 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 101 101 ZINC 2 (BY SIMILARITY).  
FT METAL 104 104 ZINC 2 (BY SIMILARITY).  
FT METAL 107 107 ZINC 2 (BY SIMILARITY).  
FT METAL 115 115 ZINC 2 (BY SIMILARITY).  
FT METAL 164 164 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 359 AA; 39137 MW; BAB1D6B3AF086DAB CRC64;

Query Match 62.1%; Score 1092; DB 1; Length 359;  
 Best Local Similarity 65.7%; Pred. No. 2.7e-82;  
 Matches 224; Conservative 27; Mismatches 56; Indels 32; Gaps 7;

QY 1 ETGATDVRFKVLYCGVCHSDIHMKNMGSTSTPIVPGHGLVGVVTEVCKVKKPK-SWR 59  
 DB 32 ETGEKDTVFKVLYCGICSHLHMKNMGSTSTPIVPGHGLVGVVTEVCKVKKPKVGR 91

QY 60 QGRCLHGRLEPTCNCIHLHLENCPNLIQYGSKYDGTMTYGGYSNNMTDDEHFI 113  
 DB 92 VGVGCIWGSRC-SCENCCTDHLHLENCPKQILYGAKYDGTMTYGGYSNNMTDDEHFI 150

QY 120 PDLNPLDGAAPLLCAGITTYSPWRYGGLDKPGHGLVGVVTEVCKVKKPK-SWR 179  
 DB 151 PDLNPLDGAAPLLCAGITTYSPWRYGGLDKPGHGLVGVVTEVCKVKKPK-SWR 194

QY 180 SLVPLIKEGG-----SVGTSP-----ALMH-----SLRTDODQMAAMSTMDGI 225  
 DB 195 HVAVKFAKAMGVKVTISTSPKKEEALHGLGADSFVSRDQDQMAAIGTMDGI 254

QY 226 AVRPLEPLISLLKXNGKVVTVGIAVQPLDLPVPLIIGRMVAGSAIGMKETQEMID 285  
 DB 255 AQHPLPLIGLNGHGLVGMVGAPEKPLELPVPLIIGRMVAGSAIGMKETQEMID 314

QY 286 AHNITADIEVIPIDYLTAMERVKDVRFRFVIDVENTL 326  
 DB 315 ARHNITADIEVIPIDYLTAMERLVKADVRYRVIDIGNTL 355

RESULT 2  
 MTD\_MESCR  
 ID MTD\_MESCR STANDARD; PRT; 361 AA.  
 AC P3257;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol  
 dehydrogenase).  
 GN EL13.  
 OS Mesembryanthemum crystallinum (Common ice plant).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllales; Aizoaceae; Mesembryanthemum.  
 OX NCBI\_TaxID=3544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Michalowski C.B., Bohnert H.J.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step  
 CC by which translocated mannitol is committed to central metabolism  
 CC and, by regulating mannitol pool size, is important in regulating  
 CC salt tolerance at the cellular level (By similarity).  
 CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC -!- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol  
 CC dehydrogenase.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U79770; A8338503.1; -;  
 CC PIR; T12571; T12571.  
 CC InterPro; IPR002328; ADH\_zinc.  
 CC InterPro; IPR002085; Adh\_zn family.  
 CC Pfam; PF00107; ADH\_zinc\_N; 1.  
 CC PROSITE; PS00059; ADH\_ZINC; 1.

KW Oxidoreductase; Zinc; Metal-binding; NAD.  
 FT METAL 51 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 73 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 104 ZINC 2 (BY SIMILARITY).  
 FT METAL 107 ZINC 2 (BY SIMILARITY).  
 FT METAL 110 ZINC 2 (BY SIMILARITY).  
 FT METAL 118 ZINC 2 (BY SIMILARITY).  
 FT METAL 167 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 361 AA; 39068 MW; 3BA2BBF946144D10 CRC64;

Query Match 55.2%; Score 971; DB 1; Length 361;  
 Best Local Similarity 55.4%; Pred. No. 2.4e-72;  
 Matches 190; Conservative 52; Mismatches 63; Indels 38; Gaps 6;

QY 2 TGATDVRFKVLYCGVCHSDIHMKNMGSTSTPIVPGHGLVGVVTEVCKVKKPK-SWR 60  
 DB 36 TGEQDTPFKVLYCGICSHLHLENCPNLIQYGSKYDGTMTYGGYSNNMTDDEHFI 95

QY 61 GRCMLHGRLEPTCNCIHLHLENCPNLIQYGSKYDGTMTYGGYSNNMTDDEHFI 120  
 DB 96 GVGCMVGSRC-SCSCENHLENCPNLIQYGSKYDGTMTYGGYSNNMTDDEHFI 154

QY 121 DNLPLDGAAPLLCAGITTYSPWRYGGLDKPGHGLVGVVTEVCKVKKPK-SWR 180  
 DB 155 DNMLDGAAPLLCAGITTYSPWRYGGLDKPGHGLVGVVTEVCKVKKPK-SWR 195

QY 181 LVPLPLIKEGGSYG-----TSPLMHS-----LLRTDQDQMAAMSTMDGI 223  
 DB 196 LHMVAVKFAKAMGVKVTISTSPKKEEALHGLGADSFVSRDQDQMAAIGTMDGI 255

QY 224 VPAVRPLEPLISLLKXNGKVVTVGIAVQPLDLPVPLIIGRMVAGSAIGMKETQEMID 283  
 DB 256 VSAHPLPLIGLNGHGLVGMVGAPEKPLELPVPLIIGRMVAGSAIGMKETQEMID 315

QY 284 PAAHNITADIEVIPIDYLTAMERVKDVRFRFVIDVENTL 326  
 DB 316 PAAHNDIKSDIEVVPMDYVNTAMERLLKGVRYRVIDVANTL 358

RESULT 3  
 MTD\_PETCR  
 ID MTD\_PETCR STANDARD; PRT; 337 AA.  
 AC P42754;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol  
 dehydrogenase) (Fragment).  
 GN EL13.  
 OS Petroselinum crispum (Parsley) (Petroselinum hortense).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Campanulids; Apiales; Apiaceae; Apioidae; apioid superclade;  
 OC Apium clade; Petroselinum.  
 OX NCBI\_TaxID=4043;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93099840; PubMed=1464303;  
 RA Kiedrowski S., Kawalleck P., Hahlbrock K., Somssich I.E., Dangel J.L.;  
 RT "Rapid activation of a novel plant defense gene is strictly dependent  
 RT on the Arabidopsis RPM1 disease resistance locus.";  
 RL EMBO J. 11:4677-4684(1992).  
 CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step  
 CC by which translocated mannitol is committed to central metabolism  
 CC and, by regulating mannitol pool size, is important in regulating  
 CC salt tolerance at the cellular level (By similarity).  
 CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

DR InterPro: IPR002328; ADH zinc.
DR InterPro: IPR002085; Adh_zn family.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL 46 46 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 68 68 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 102 102 ZINC 2 (BY SIMILARITY).
FT METAL 105 105 ZINC 2 (BY SIMILARITY).
FT METAL 113 113 ZINC 2 (BY SIMILARITY).
FT METAL 162 162 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 6 6 E -> Q (IN REF. 1).
FT CONFLICT 17 17 E -> N (IN REF. 1).
FT CONFLICT 20 20 I -> V (IN REF. 1).
SQ SEQUENCE 357 AA; 38245 MW; 2C27B3C2BF030166 CRC64;

Query Match 54.6%; Score 960; DB 1; Length 357;
Best Local Similarity 55.4%; Pred. No. 1.9e-71;
Matches 190; Conservative 47; Mismatches 68; Indels 38; Gaps 5;

QY 2 TGATDVRFKVLGVGCHSDIIMAXNDWGTSTPIVPGHVLGVVTEVGVCKKPKSW-RQ 60
DB 31 TGEKDVRFKVLGVGCHSDIIMAXNDWGTSTPIVPGHVLGVVTEVGVCKKPKSW-RQ 60
QY 61 GRCWLHGLRPTCEHCILHLENYCNLIQTGYSGKYDGTMTYGGYSNNKVTDEHFVIRIP 120
DB 91 GVGVMAGSCR-SCDSCNDGDENYCPKMLITSGAKNFDFTWTHGGYSDHVMCAEDFIIRIP 149
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGHMLGVHVEPRPRPPLNLPOLWSRLQS 180
DB 150 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGHMLGVHVEPRPRPPLNLPOLWSRLQS 180
QY 181 LVPPLIKEGSGYGTSPALMHS-----LLRTDDQDQAAAMSTMDGIIDT 223
DB 191 LCHVAVFPAKANGTKVTVISERKRDVAFTLGDADFVSRDPKMDAMGTMDGIIDT 250
QY 224 VPAVRPLEPLSLKNGKVVTVGVIAVQPLDVPFPLIIGRMVAGSAGIGMKETQEMID 283
DB 251 VSATHPL-PLLLGKLNKGLKLVGAPAPLEPLVPFPLIFGRKVVGVSGVIGKETQEMVD 310
QY 284 FAEENHTADIEVPIIDVINTAMERVKVQVRFVIVENTL 326
DB 311 LAGKENITADIELISADYVNTAMERLAKADVKYRFVIDVANTM 353

RESULT 5
ID WT22_ATH STANDARD; PRT; 359 AA.
AC Q02972;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable mannitol dehydrogenase 2 (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase 2).
GN ELI3-2 OR AT4G37990 OR F20D10.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=93099840; PubMed=1464303;
RA Kiedrowski S., Kawalleck P., Hahlbrock K., Somsich I.E., Dangl J.L.;
RT "Rapid activation of a novel plant defense gene is strictly dependent on the Arabidopsis RPM1 disease resistance locus.";
RL EMOB J. 11:4677-4684(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;

```

```

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Voickaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoerge W., Brandt P., Grivelli L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier S., Mache R., Mueller M.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Waltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzreger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer B., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Borkova D., Bloeker H., Schafie M., Benes V., Rechmann S.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Loehner T.-H.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedors T., Weber N., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney J., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.";
RL Nature 402:769-777(1999).
CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step
CC by which translocated mannitol is committed to central metabolism
CC and, by regulating mannitol pool size, is important in regulating
CC salt tolerance at the cellular level (By similarity).
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X67815; CAB48026.1; -
DR EMBL; AL035538; CAB37539.1; -
DR EMBL; ALJ61592; CAB80464.1; -
DR F01; S28343; S28043.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn family.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL 46 46 ZINC 1 (CATALYTIC) (BY SIMILARITY).

```

FT METAL 68 68 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 99 99 ZINC 2 (BY SIMILARITY).  
FT METAL 102 102 ZINC 2 (BY SIMILARITY).  
FT METAL 105 105 ZINC 2 (BY SIMILARITY).  
FT METAL 113 113 ZINC 2 (BY SIMILARITY).  
FT METAL 162 162 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 359 AA; 38942 MW; B691F988AD4842A5 CRC64;  
  
Query Match 54.6%; Score 959; DB 1; Length 359;  
Best Local Similarity 55.1%; Pred. No. 2.3e-71;  
Matches 193; Conservative 39; Mismatches 68; Indels 50; Gaps 5;  
  
QY 1 ETGATDVRKVLVGVCHSDIHMKNKNDWGTSTPIVPGHELGVVTVGCKVKFKSWRQ 60  
DB 30 ETGEKDVRFKVLFCGCHSDIHMKNKNDWGTSTPIVPGHELGVVTVGCKVKFKSWRQ 89  
QY 61 -----GRCWLHGRLEPFCNCIHLNENYCPNLIQTYGSKYDGTWYGYNNMVYDE 113  
DB 90 VGVGCLVSSC-----GSCDSCTEGMENYCPKSIQTYGFPYDNTITYGYSDHVMVCEE 142  
QY 114 HPTVIRPDNPLDGAAPLLCAGITTYSPWRYGLDKPGMLGVNPPRRSRPPLNPLGL 173  
DB 143 GFVIRPDNPLDGAAPLLCAGITTYSPWRYGLDKPGMLGVNPPRRSRPPLNPLGL 188  
QY 174 WGSRLQSLVPLPKEGSSYGTSPALMHS-----LLRTDQDQMEAAAMST 216  
DB 189 GG-----LGHVGVKFAKAMGKVTVIISTSEKKRDEANRLGADAFVSRDPKQIKDAMGT 243  
QY 217 MDGIIDTPVAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGMK 276  
DB 244 MDGIIDTPVAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGMK 303  
QY 277 ETOEMIDFAEHNTIADIEVPIIDYNTAMERVVKVORFRFVIDVENTL 326  
DB 304 ETOEMIDMAGKNITADIELISADYNTAMERLEKADVRVRFVIDVANTL 353

RESULT 6  
MTD APIGR STANDARD; PRT; 365 AA.  
AC Q38707; O82461;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol  
dehydrogenase).  
GN MTD.  
OS Apium graveolens (Celery).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC Campanulids; Apiales; Apiaceae; Apioidae; apioid superclade;  
OC Apium clade; Apium.  
OX NCBI\_TaxID=4045;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95365325; PubMed=7638158;  
RA Williamson J.D., Stoop J.M.H., Massel M.O., Conkling M.A., Pharr D.M.;  
RT "Sequence analysis of a mannitol dehydrogenase cDNA from plants  
reveals a function for the pathogenesis-related protein EL13.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:7148-7152 (1995).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Williamson J.D., Guo W.-W., Pharr D.M.;  
RT "Cloning and characterization of a genomic clone encoding mannitol  
dehydrogenase from celery (Apium graveolens).";  
RL (In) Plant Gene Register PGR98-137.  
RN [3]  
RN CHARACTERIZATION  
RA Stoop J.M.H., Chilton W.S., Pharr D.M.;  
RT "Substrate specificity of the NAD-dependent mannitol dehydrogenase  
from celery.";  
RL Phytochemistry 43:1145-1150 (1996).  
CC -i- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step

CC by which translocated mannitol is committed to central metabolism  
and, by regulating mannitol pool size, is important in regulating  
salt tolerance at the cellular level.  
CC -i- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.  
CC -i- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
CC -i- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -i- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
CC  
CC EMBL; U24561; AAC15467.1; -;  
CC EMBL; AF067082; AAC61854.1; -;  
CC InterPro; IPR002328; ADH\_zinc.  
CC InterPro; IPR002085; Adh\_zn\_family.  
CC InterPro; IPR000205; NAD\_BS.  
CC InterPro; IPR000051; SAM\_Bind.  
CC Pfam; PF00107; ADH\_zinc\_N; 1.  
CC PROSITE; PS00059; ADH\_ZINC; 1.  
CC Oxidoreductase; Zinc; Metal-binding; NAD.  
KW METAL 50 50 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 72 72 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT METAL 103 103 ZINC 2 (BY SIMILARITY).  
FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
FT METAL 109 109 ZINC 2 (BY SIMILARITY).  
FT METAL 117 117 ZINC 2 (BY SIMILARITY).  
FT METAL 166 166 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
FT CONFLICT 47 47 C -> S (IN REF. 2).  
FT CONFLICT 120 120 T -> I (IN REF. 2).  
SQ SEQUENCE 365 AA; 39690 MW; A46E394F5DCB077F CRC64;  
  
Query Match 53.8%; Score 946; DB 1; Length 365;  
Best Local Similarity 56.5%; Pred. No. 2.7e-70;  
Matches 192; Conservative 42; Mismatches 74; Indels 32; Gaps 5;  
  
QY 2 TGATDVRKVLVGVCHSDIHMKNKNDWGTSTPIVPGHELGVVTVGCKVKKEK-SWRQ 60  
DB 35 TGEKDVRLKVLFCGCHSDIHMKNKNDWGTSTPIVPGHELGVVTVGCKVKKEKGVNV 94  
QY 61 GRWLHGRLEPFCNCIHLNENYCPNLIQTYGSKYDGTWYGYNNMVTDHPIVRIP 120  
DB 95 GIGLVGSCR-SCBSCDNRSHCHENTIDTYGSIYFDGTWCHGGYSTMTWADHFIILRW 153  
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMLGVNPPRRSRPPLNPLGLSGSRLOS 180  
DB 154 KNPLDGAAPLLCAGITTYSPWRYGLDKPGMLGVNPPRRSRPPLNPLGLSGSRLOS 197  
QY 181 LVPLPKEGSSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDTPVA 226  
DB 198 VAVTKAFAGQVTVIDISESKRKEALEKLGADFLNSDQEMKGAARSLDGIIDTPV 257  
QY 227 VRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGMKETQEMIDFAA 286  
DB 258 NHPLAPLPDLKPKNGKLVWGAPEKPELPVPLSKKRLKGLGTTGGIKETQEMIDFAA 317  
QY 287 EHNITADIEVPIIDYNTAMERVVKVORFRFVIDVENTL 326  
DB 318 KHNITADIEVPIIDYNTAMERVVKVORFRFVIDDIANTM 357  
  
RESULT 7  
MTD ARATH  
ID MTD ARATH STANDARD; PRT; 360 AA.  
AC P42734;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)



OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
 OX NCBI\_TaxID=3879;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=cv. Siriver; TISSUE=Stem;  
 RX MEDLINE=20044095; PubMed=10579494;  
 RA Brill E.M., Abrahams S., Hayes C.M., Jenkins C.L., Watson J.M.;  
 RT "Molecular characterization and expression of a wound-inducible cDNA  
 encoding a novel cinnamyl-alcohol dehydrogenase enzyme in lucerne  
 (Medicago sativa L.)";  
 RL Plant Mol. Biol. 41:279-291(1999).  
 CC -1- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step  
 by which translocated mannitol is committed to central metabolism  
 and, by regulating mannitol pool size, is important in regulating  
 salt tolerance at the cellular level (By similarity).  
 CC -1- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.  
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF083333; AAC35846.1; -;  
 DR InterPro; IPR002328; ADH\_zinc.  
 DR Pfam; PF00107; ADH\_zinc\_N; 1.  
 DR PROSITE; PS00059; ADH\_ZINC; 1.  
 KW Oxidoreductase; Zinc; Metal-binding; NAD.  
 FT METAL 50 50 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 72 72 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 103 103 ZINC 2 (BY SIMILARITY).  
 FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
 FT METAL 109 109 ZINC 2 (BY SIMILARITY).  
 FT METAL 117 117 ZINC 2 (BY SIMILARITY).  
 FT METAL 165 165 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 359 AA; 39027 MW; D84E383E0F195A1 CRC64;  
 Query Match 49.5%; Score 870.5; DB 1; Length 359;  
 Best Local Similarity 52.0%; Pred. No. 4.2e-64;  
 Matches 178; Conservative 42; Mismatches 85; Indels 37; Gaps 5;  
 QY 1 ETGATDVRPKVLYGCVCHSDIHMKNWGTSTYTPVPGHELGVVTVGCKVKPKSWRQ 60  
 DB 34 ENGDDVSVKILYGVCHSDLTLEKNWGTFTYTPVPGHELGVVTVGKINVKKFRVGDN 93  
 QY 61 GRCLWHLGRLEPTCNCIHLNENYCNLIQTYGSKYVDGTYGYSNNMVTDEHFIVIP 120  
 DB 94 VGVGVIVRESQTCNCNQLDLEQCEKFPYFTNSP-YKGRTRYGYSDFVHVHQRYVWQFP 152  
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPCGHLGVWPRFRSRPPLNPLGLWCSRLQS 180  
 DB 153 DNLPLDGAAPLLCAGITTYSPWRYGLDKPCGHLGVWPRFRSRPPLNPLGLWCSRLQS 193  
 QY 181 LVPPLEIKEGGSYG-----TSP-----ALMHSLLRTDODQBEAAMTMDGIIDT 223  
 DB 194 LGHVAIKFGKAFGLKVTVISPNKTEAIDKLGADSFVSKPEKMAAGTMDYIIDT 253  
 QY 224 VPAVRPLEPLSLKTNKGVTVGVAGVLPDLPFPLLIIGRWAGSAIGCKKETQEMID 283  
 DB 254 ISAAHSLAPLLGLKLNGLKVTGVLSPKPLSLVFPPLVAGRLIGGSGNIGCKKETQEMLD 313  
 QY 284 FAAGHNTIADLEVPIDVINTAMERVVKVKDVRFRFVIDVENT 325  
 DB 314 FCGKHNTIADLEIYNHINTEAMERLKHADVKYRFVIDVANS 355

MTD3 STYHU  
 ID MTD3 STYHU STANDARD; PRT; 363 AA.  
 AC Q43138; 1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, last sequence update)  
 DT 10-OCT-2003 (Rel. 42, last annotation update)  
 DE Probable mannitol dehydrogenase 3 (EC 1.1.1.255) (NAD-dependent  
 mannitol dehydrogenase 3).  
 GN CAD3.  
 OS Stylosanthes humilis (Townselle stylo).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;  
 OC Stylosanthes.  
 OX NCBI\_TaxID=35628;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Paterson; TISSUE=Stem;  
 RA Nourse J.P., Manners J.M., Curtis M.D., Abrahams S.L., Watson J.M.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step  
 by which translocated mannitol is committed to central metabolism  
 and, by regulating mannitol pool size, is important in regulating  
 salt tolerance at the cellular level (By similarity).  
 CC -1- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.  
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 family.  
 CC -1- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol  
 dehydrogenase.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; L36456; AAA74883.1; -;  
 DR InterPro; IPR002328; ADH\_zinc.  
 DR Pfam; PF00107; ADH\_zinc\_N; 1.  
 DR PROSITE; PS00059; ADH\_ZINC; 1.  
 KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.  
 FT METAL 51 51 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 73 73 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 104 104 ZINC 2 (BY SIMILARITY).  
 FT METAL 107 107 ZINC 2 (BY SIMILARITY).  
 FT METAL 110 110 ZINC 2 (BY SIMILARITY).  
 FT METAL 118 118 ZINC 2 (BY SIMILARITY).  
 FT METAL 168 168 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 363 AA; 39551 MW; E8EB217274513D90 CRC64;  
 Query Match 48.9%; Score 860.5; DB 1; Length 363;  
 Best Local Similarity 50.3%; Pred. No. 2.8e-63;  
 Matches 172; Conservative 52; Mismatches 85; Indels 33; Gaps 6;  
 QY 1 ETGATDVRPKVLYGCVCHSDIHMKNWGTSTYTPVPGHELGVVTVGCKVKPKSWRQ 60  
 DB 35 DIGEDVALEVLVYIGICHTDLFMAKNDFGNSIYPVPGHGVIGIVAEVGSVKYKVGDK 94  
 QY 61 GRCLWHLGRLEPTCNCIHLNENYCNLIQTYGSKYVDGTYGYSNNMVTDEHFIVIP 120  
 DB 95 VGVGVIVRESQTCNCNQLDLEQCEKFPYFTNSP-YKGRTRYGYSDFVHVHQRYVWQFP 154  
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPCGHLGVWPRFRSRPPLNPLGLWCSRLQ 179  
 DB 155 BGLPLDGGSSSLCWGY-SHSPKLYTGLDKPGLHGV-----VGLGG--LG 197  
 QY 180 SLVPLPIKEGG-----SYGTSPALMES-----LLRTDQDQBEAAMTMDGIIDT 224  
 DB 198 HMVAKFAKTHGLKITVISTSPPTKKEEAIXNLGADSFVSRDPDQMEAPKETLGLIIDTV 257



FT METAL 114 114 ZINC 2 (BY SIMILARITY).  
 SQ SEQUENCE 357 AA; 38834 MW; 56P8133B9B3097FC CRC64;  
 Query Match 42.1%; Score 740; DB 1; Length 357;  
 Best Local Similarity 42.7%; Pred. No. 2.1e-53;  
 Matches 147; Conservative 61; Mismatches 86; Indels 50; Gaps 5;

QY 3 GATDVRFKVLYCGVCHSDIHMAKNDGTSTYPIVRGHELVGVTEVCGCKVKKESMRQ-- 60  
 DB 33 GPEDVIRVYICIGICHSDDLVQWENEMGMSYPMWPGHEVGVVTEIGSEVKKFKVGEHVG 92  
 QY 61 -----GRCMLHGLRLPTCNCIHLNENYCPNLITQYGSKYDGTMTYGGYNNMVDDEHF 115  
 DB 93 VGCIVGSC-----RSCSNCNGSMGQYCSKRIWTYNDVNHDTPTQGGFASWVVDQMF 145  
 QY 116 IVRIPNPLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPLNPLGLWG 175  
 DB 146 VVRIPENPLEQAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPLNPLGLWG 194  
 QY 176 SRLQSLVPLLIKGGSGYGTSPALMHS-----LLRTDQDQMEAAKSTMD 218  
 DB 195 MG-----VKIAKAFGLHVTVISSDKKKERALEVLGADAYLVSKDAEKMQEASLD 246  
 QY 219 GIIDTVPVRPLEPLSLKTKNGKVTYVGIAGVQPLDLPVPLIIGRMVAGSAIGMKET 278  
 DB 247 YIMDTIPVAHPLEPLALTKNGKLVMLGVVPEPLHFTVPLLLILGRSISAGSFIGSMEET 306  
 QY 279 QEMIDFAAHNITADIEVIDYNTAMERLVKVDVRFVVDV 322  
 DB 307 QETLDFCAKKVSSMIEVGLDYINTAMERLVKVDVRFVVDV 350

RESULT 12  
 CAD7 PICAB STANDARD; PRT; 357 AA.  
 ID CAD7 PICAB  
 AC Q08350;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cinnamyl-alcohol dehydrogenase 7/8 (EC 1.1.1.195) (CAD 7/8).  
 GN CAD7 AND CAD8.  
 OS Picea abies (Norway spruce) (Picea excelsa).  
 OC Bukayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.  
 OX NCBI\_TaxID=3329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94033286; PubMed=8219046;  
 RA Galliano H., Cabane M., Ekerskorn C., Lottspeich F.,  
 RA Sandermann H. Jr., Ernst D.;  
 RT "Molecular cloning, sequence analysis and elicitor/ozone-induced  
 accumulation of cinnamyl alcohol dehydrogenase from Norway spruce  
 (Picea abies L.)."  
 RT Plant Mol. Biol. 23:145-156(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Schubert R., Sperisen C., Mueller-Starck G., La Scala S., Ernst D.,  
 RA Sandermann H. Jr., Haeger K.-P.;  
 RT "The cinnamyl alcohol dehydrogenase gene family in Picea abies (L.)  
 Karst.: genomic sequences, Southern hybridization, genetic analysis  
 and phylogenetic relationships."  
 RT Trees 12:453-463(1998).  
 CC -!- FUNCTION: This protein catalyzes the final step in a branch of  
 CC phenylpropanoid synthesis specific for production of lignin  
 CC monomers. It acts on coniferyl-, sinapyl-, 4-coumaryl- and  
 CC cinnamyl-alcohol.  
 CC CATALYTIC ACTIVITY: Cinnamyl alcohol + NADP(+) = cinnamaldehyde +  
 CC NADPH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- PATHWAY: Lignin biosynthesis.  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X72675; CAA51226.1; --  
 DR EMBL; AJ001925; CAA05096.1; --  
 DR EMBL; AJ001926; CAA05097.1; --  
 DR PIR; S39509; S39509.  
 DR InterPro; IPR002328; ADH\_zinc.  
 DR Pfam; PF00107; ADH\_zinc\_N; I.  
 DR PROSITE; PS00059; ADH\_ZINC; I.  
 KW Oxidoreductase; Zinc; Metal-binding; NADP; Lignin biosynthesis;  
 KW Multigene family.  
 FT METAL 47 47 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 69 69 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 103 100 ZINC 2 (BY SIMILARITY).  
 FT METAL 103 103 ZINC 2 (BY SIMILARITY).  
 FT METAL 106 106 ZINC 2 (BY SIMILARITY).  
 FT METAL 114 114 ZINC 2 (BY SIMILARITY).  
 FT METAL 163 163 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 357 AA; 38777 MW; 8CD26BC41B87CA92 CRC64;  
 Query Match 42.1%; Score 740; DB 1; Length 357;  
 Best Local Similarity 42.7%; Pred. No. 2.1e-53;  
 Matches 147; Conservative 61; Mismatches 86; Indels 50; Gaps 5;

QY 3 GATDVRFKVLYCGVCHSDIHMAKNDGTSTYPIVRGHELVGVTEVCGCKVKKESMRQ-- 60  
 DB 33 GPEDVIRVYICIGICHSDDLVQWENEMGMSYPMWPGHEVGVVTEIGSEVKKFKVGEHVG 92  
 QY 61 -----GRCMLHGLRLPTCNCIHLNENYCPNLITQYGSKYDGTMTYGGYNNMVDDEHF 115  
 DB 93 VGCIVGSC-----RSCSNCNGSMGQYCSKRIWTYNDVNHDTPTQGGFASWVVDQMF 145  
 QY 116 IVRIPNPLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPLNPLGLWG 175  
 DB 146 VVRIPENPLEQAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPLNPLGLWG 194  
 QY 176 SRLQSLVPLLIKGGSGYGTSPALMHS-----LLRTDQDQMEAAKSTMD 218  
 DB 195 MG-----VKIAKAFGLHVTVISSDKKKERALEVLGADAYLVSKDAEKMQEASLD 246  
 QY 219 GIIDTVPVRPLEPLSLKTKNGKVTYVGIAGVQPLDLPVPLIIGRMVAGSAIGMKET 278  
 DB 247 YIMDTIPVAHPLEPLALTKNGKLVMLGVVPEPLHFTVPLLLILGRSISAGSFIGSMEET 306  
 QY 279 QEMIDFAAHNITADIEVIDYNTAMERLVKVDVRFVVDV 322  
 DB 307 QETLDFCAKKVSSMIEVGLDYINTAMERLVKVDVRFVVDV 350

RESULT 13  
 CADH\_PINTA STANDARD; PRT; 357 AA.  
 ID CADH\_PINTA  
 AC P41637;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) (CAD).  
 OS Pinus taeda (Loblolly pine).  
 OC Bukayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 OX NCBI\_TaxID=3352;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Xylem;  
 RX MEDLINE=95327049; PubMed=7603432;



```

QY 61 -----GRCWLHGLRLPTCENCIIHLENYCPNLIQTYGSKYDGTMTYGGYSNNMVTDEHF 115
Db 93 VGCIVGSC-----RSCGNCNQSQEQYCSKRIWTYNDVNDGDTPTQGGFASWVVDQMF 145
QY 116 IVRIPNLPLDGAAPLLCAGITTYSPWRYYGLDKFGMHLGVEMPRFRSRPLNLPGLWG 175
Db 146 VVRIPENLPLDGAAPLLCAGITTYSPWRYYGLDKFGMHLGVEMPRFRSRPLNLPGLWG 194
QY 176 SRLQSLVPLPLKEGSGSYTSPALMHS-----LLRTQDQDQEAAMSTMD 218
Db 195 MG-----VKIAKAFGLHVTVISSDKKKEAMEVLGADAYLVSKDTCKMERAESLD 246
QY 219 GIIDTPAVRPLEPLISLKTNGKVVTVGIAVQPLDLVPFLLITGRKMVAGSAIGMKET 278
Db 247 YIMDTIPVAHLEPLLEALLKNGKLMVGVPELHFVTPLLIGRRSIAGSFGSMEET 306
QY 279 QEMIDFAAENHITADIEVIPIDYLTAMERVVKDVRFRFVIDV 322
Db 307 QETLDFCAEKVKSSMIEVVGLDYINTAMERLEKNDVRYRFVVDV 350

RESULT 15
CADI_ARACO
ID CADI_ARACO STANDARD; PRT; 360 AA.
AC P42495;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cinnamyl-alcohol dehydrogenase 1 (EC 1.1.1.195) (CAD).
GN CADI.
OS Aralia cordata (Udo) (Cordate spikenard).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Apiales; Araliaceae; Aralia.
OX NCBI_TaxID=29746;
[1]
RN SEQUENCE FROM N.A.
RA Hibino T., Shibata D., Chen J.-O., Higuchi T.;
RT "Cinnamyl alcohol dehydrogenase from Aralia cordata: cloning of the
RT cDNA and expression of the gene in lignified tissues.";
RL Plant Cell Physiol. 34:659-665(1993).
CC -1- FUNCTION: This protein catalyzes the final step in a branch of
CC phenylpropanoid synthesis specific for production of lignin
CC monomers. It acts on coniferyl-, sinapyl-, 4-coumaryl- and
CC cinnamyl-alcohol.
CC -1- CATALYTIC ACTIVITY: Cinnamyl alcohol + NADP(+) = cinnamaldehyde +
CC NADPH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Lignin biosynthesis.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13991; BAA03099.1; -.
CC InterPro; IPR002328; ADH_zinc.
CC InterPro; IPR002085; Adh_zn_family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NADP; Lignin biosynthesis.
FT METAL 47 47 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 69 69 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 100 100 ZINC 2 (BY SIMILARITY).
FT METAL 103 103 ZINC 2 (BY SIMILARITY).
FT METAL 106 106 ZINC 2 (BY SIMILARITY).
FT METAL 114 114 ZINC 2 (BY SIMILARITY).
FT METAL 163 163 ZINC 1 (CATALYTIC) (BY SIMILARITY).

```

```

SQ SEQUENCE 360 AA; 39129 MW; 28F1D980E08C4096 CRC64;
Query Match 39.3%; Score 691.5; DB 1; Length 360;
Best Local Similarity 44.0%; Pred. No. 2.1e-49;
Matches 150; Conservative 52; Mismatches 100; Indels 39; Gaps 9;
QY 1 ETGATDVRFKVLYCGVCHSDIHMAKNKMGTSYPIVPGHLYGVVTEUCCKYKKEKSWQ 60
Db 31 ETGPEDEVFKIYICGICHTDIHQTKNDLGASNYPMVPGHEVVGEVGSQVTKPK- -V 87
QY 61 GRCWLHGLRL-----RPTCENCIIHLENYCPNLIQTYGSKYDGTMTYGGYSNNMVTDEHF 117
Db 88 GDCVGDGTIVGCKTCRCADVEQYCNKKIWSYNDVTDGKPTQGGFSGHVVVDQKFW 147
QY 118 RIPNLPDGAAPLLCAGITTYSPWRYYGLDK-FGMHLGVEMPRFRSRPLNLPGLMGS 176
Db 148 KIPDGAPEQAAPLLCAGITTYSPWRYYGLDK-FGMHLGVEMPRFRSRPLNLPGLMGS 196
QY 177 RLQSLVPLPLKEGSGSYGT-----SPALNH-----SLARTDQDQEAAMSTMDGIID 222
Db 197 GVK-----LAKAMGHVTVISSDKKKEEAIDHLGADAYLVSSDATQCEAADSLOYIID 251
QY 223 TVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLVPFLLI-ICRMVAGSAIGMKETQM 281
Db 252 TVPVFHPLEPVLSLKLDGKLLMGVINTPLQF-ISPVMVLMGRKAITGSGFSGMKEEEM 310
QY 282 IDFAAENHITADIEVIPIDYLTAMERVVKDVRFRFVIDV 322
Db 311 LDFCNKGITSTIEVWQMDYINTAFERLEKNDVRYRFVVDV 351

```

Search completed: July 6, 2004, 13:37:17

Job time : 9.38046 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:14 ; Search time 31.8458 Seconds  
(without alignments)  
3229.910 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKLYGCVCHSD.....ERVVKDVRFRVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1064	60.5	359	10	Q9ATW1	Q9atw1 fragaria an
2	984	56.0	362	10	Q94G59	Q94g59 populus tre
3	970	55.2	363	10	Q7XAB2	Q7xab2 camptotheca
4	950.5	54.1	359	10	Q8H018	Q8h018 solanum tub
5	936	53.2	362	10	Q8L7U8	Q8l7u8 populus tre
6	881	50.1	360	10	Q94K02	Q94k02 arabidopsis
7	876	49.8	360	10	Q8LB84	Q8lb84 arabidopsis
8	853	48.5	363	10	Q65621	Q65621 arabidopsis
9	839	47.7	370	10	Q8S411	Q8s411 lolium pere
10	819.5	46.6	375	10	Q8S410	Q8s410 arabidopsis
11	802	45.6	376	10	Q9SJ25	Q9sj25 arabidopsis
12	790	44.9	420	10	Q7X898	Q7x898 oryza sativ
13	784.5	44.6	289	10	Q9M722	Q9m722 lycopersico
14	755.5	43.0	360	10	Q7XW00	Q7xw00 oryza sativ
15	712.5	40.5	391	10	Q8H809	Q8h809 oryza sativ
16	707	40.2	410	10	Q7XLD5	Q7xld5 oryza sativ

17	679	38.6	357	10	Q9M632	Q9m632 populus tre
18	675	38.4	357	10	Q9FSC7	Q9fsc7 populus tri
19	666	37.9	407	10	Q8S412	Q8s412 lolium pere
20	662	37.7	361	10	Q947S1	Q947s1 festuca aru
21	660	37.5	361	10	Q947S3	Q947s3 festuca aru
22	660	37.5	361	10	Q947S2	Q947s2 festuca aru
23	654	37.2	361	10	Q947S0	Q947s0 festuca aru
24	648	36.9	356	10	Q9FUN8	Q9fun8 eucalyptus
25	641	36.5	335	10	Q8W420	Q8w420 medicago sa
26	626.5	35.6	348	2	Q9AE96	Q9ae96 mycobacteri
27	622.5	35.4	350	16	Q88K65	Q88k65 pseudomonas
28	622.5	35.4	350	16	Q884B3	Q884b3 pseudomonas
29	622	35.4	349	2	Q8KRC3	Q8krc3 myxococcus
30	621.5	35.4	352	5	Q9ULF0	Q9ulf0 leishmania
31	621	35.3	347	16	Q92MD4	Q92md4 rhizobium m
32	616.5	35.1	355	16	Q8XQ04	Q8xq04 raietonia s
33	614	34.9	365	10	Q8L9U1	Q8l9ul arabidopsis
34	613.5	34.9	347	16	Q82I44	Q82i44 streptomyce
35	613.5	34.9	362	16	Q9CBQ3	Q9cbq3 mycobacteri
36	612	34.8	321	10	Q04079	Q04079 zinnia eleg
37	610.5	34.7	354	10	Q8H859	Q8h859 oryza sativ
38	609	34.6	368	16	Q8UF43	Q8uf43 agrobacteri
39	596.5	33.9	349	16	Q06007	Q06007 bacillus su
40	596	33.9	358	16	Q882D1	Q882d1 pseudomonas
41	588.5	33.5	346	16	Q8ES57	Q8es57 oceanobacil
42	587.5	33.4	355	10	Q8CAI3	Q8cai3 arabidopsis
43	583.5	33.2	352	16	Q8PRD2	Q8prd2 xanthomonas
44	581.5	33.1	353	16	Q91I09	Q91i19 pseudomonas
45	580.5	33.0	352	16	Q8YI07	Q8yi07 raietonia s

## ALIGNMENTS

### RESULT 1

Q9ATW1 PRELIMINARY; PRT; 359 AA.  
ID Q9ATW1  
AC Q9ATW1;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Cinnamyl alcohol dehydrogenase.  
OS Fragaria ananassa (Strawberry).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC euroids; I; Rosales; Rosaceae; Rosoideae; Fragaria.  
OX NCBI\_TaxID=3747;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Chandler;  
RA Blanco-Portales R.R., Caballero-Capullo J.J., Munoz-Blanco J.J.;  
RT "Cloning, expression and immunolocalization pattern of a cinnamyl  
RT alcohol dehydrogenase gene from strawberry (Fragaria x ananassa c.v.  
RT Chandler).";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- COFACTOR: ZINC (BY SIMILARITY).  
DR EMBL; AF220110; AAK38509.1;  
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.  
DR GO; GO:0005489; P:electron transporter activity; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0008270; P:zinc ion binding; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002328; Adh\_zinc.  
DR InterPro; IPR002085; Adh\_zn family.  
DR InterPro; IPR00345; CytC\_heme\_BS.  
DR Pfam; PF00107; Adh\_zinc\_N; 1.  
DR PROSITE; PS00059; Adh\_ZINC; 1.  
DR PROSITE; PS00130; CYTOCHROME\_C; 1.  
KW Metal-Binding; Oxidoreductase; Zinc.  
SQ SEQUENCE 359 AA; 38907 MW; 5452F249A8ECAA66 CRC64;

Query Match 60.5%; Score 1064; DB 10; Length 359;  
Best Local Similarity 64.5%; Pred. No. 6.1e-89;



```

AC Q8H0L8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alcohol NADP+ oxidoreductase.
GN DRD-1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bintje;
RA Montesano M., Hyttinen H., Wettstein R., Palva T.E.;
RT "A novel potato defense-related alcohol:NADP+ oxidoreductase induced
in response to Erwinia carotovora."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ439933; CAD29291.1; -
DR GO: GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn family.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
DR PROSITE: PS00190; CYTOCHROME C; 1.
SQ SEQUENCE 359 AA; 39388 MW; AA6B32922E2B73A66 CRC64;

Query Match 54.1%; Score 950.5; DB 10; Length 359;
Best Local Similarity 54.2%; Pred. No. 1.5e-78;
Matches 186; Conservative 56; Mismatches 62; Indels 39; Gaps 6;

QY 2 TGAATDVRFKLYCGVCHSDIHMANKNDWGTSTYPIVPGHLELVGVTVGVCKVKKFK-SWRQ 60
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 35 TGEKHQVQKVMYCGIChSDHLQKNEWNTKPMVPGHVGWVGVVGVGKVEKVGDKV 94
QY 61 GRCWLHGRLEPTCENCIHLENCNLIQTYGKYDGTMTYGYGYSNNMVTDEHFIVRIP 120
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 95 GVGCMVGSCE-KCENCTVDLENYCPRIPTYNGYSLDGLTFGYSNNMVSDEHFVVRP 153
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWPRFRSRPPLNLPGLWGSRLQS 180
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 154 ENLSMO-AAELLCAGITTYSPKTYFLGDKPMHGLGV- - - - -VGLGG----- 193
QY 181 LVPLLEKGGSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDT 223
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 194 LGHMAVKFAKAFGKTVISTSAKKQEAERLGAOSFLISRDPBQKMAAMTLDGIIDT 253
QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPFLLIIGRKVAGSAIGMKETQEMID 283
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 254 VSAVHPILPLLMKSHGKLWVGAPEKPVLELPVFPFLLMGRKLVAGSCIGMKETQEMLD 313
QY 284 EAEHNITADIEVIPDYLTAMERVVKDVRFRFVDVENTL 326
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 314 FRAGHNITPDIEVPMYDVTNRLRLKSDVKRYFVLDIGNTL 356

RESULT 5
Q8L7U8
ID Q8L7U8 PRELIMINARY; PRT; 362 AA.
AC Q8L7U8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Putative sinapyl alcohol dehydrogenase.
OS Populus tremula x Populus tremuloides.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Malpighiales; Salicaceae; Salix; Populus.
OX NCBI_TaxID=47664;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Israelsson M., Eriksson M.E., Hertzberg M., Aspeberg H., Nilsson P.,
RA Moritz T.;
RT "Changes in gene transcription in the wood-forming tissue of
transgenic hybrid aspen with increased secondary growth."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY126444; AN955578.1; -
DR GO: GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn family.
DR InterPro: IPR000345; CytC_heme_BS.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
DR PROSITE: PS00190; CYTOCHROME C; 1.
SQ SEQUENCE 362 AA; 39527 MW; BA512D2CEFC97EC6 CRC64;

Query Match 53.2%; Score 936; DB 10; Length 362;
Best Local Similarity 52.2%; Pred. No. 3.3e-77;
Matches 179; Conservative 54; Mismatches 74; Indels 36; Gaps 3;

QY 1 ETGATDVRFKLYCGVCHSDIHMANKNDWGTSTYPIVPGHLELVGVTVGVCKVKKFKSWRQ 60
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 34 DNGVEDVTIKLYCGVCHSDIHAANKNEWNTKPMVPGHVGWVGVVGVGKVEKVDQ 93
QY 61 GRCWLHGRLEPTCENCIHLENCNLIQTYGKYDGTMTYGYGYSNNMVTDEHFIVRIP 120
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 94 VGVGMVNSCKSCBYCDQSENCFKMTFTYNAQNYDGTMTYGYGYSNNMVTDEHFIVRIP 153
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWPRFRSRPPLNLPGLWGSRLQS 180
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 154 DSMPADGAPLLCAGITTYSPKTYFLGDKPMHGLGV- - - - -VGLGG----- 194
QY 181 LVPLLEKGGSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDT 223
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 195 LGHMAVKFAKAFGKTVISTSAKKQEAERLGAOSFLISRDPBQKMAAMTLDGIIDT 254
QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPFLLIIGRKVAGSAIGMKETQEMID 283
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 255 VSAVHALPLLSLKTNGKLVTLGLPEKLELPFPLVGLGRKLVAGSDIGMKETQEMLD 314
QY 284 EAEHNITADIEVIPDYLTAMERVVKDVRFRFVDVENTL 326
DQ |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 315 FCAGHNITADIEVPMYDVTNRLRLKSDVKRYFVLDIVANSL 357

RESULT 6
Q94K02
ID Q94K02 PRELIMINARY; PRT; 360 AA.
AC Q94K02;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cinnamyl-alcohol dehydrogenase CAD1 (Putative alcohol dehydrogenase)
DE (EC 1.1.1.195).
GN T22F8.230 OR CAD9 OR AT4G39330.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Bach J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinozaki K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volkovskiy S.N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation.";  
RL Genome Biol. 0:0-0(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DDJB databases.  
DR EMBL: AY087363; F:alcohol dehydrogenase activity, zinc-dependent; IEA.  
DR GO:GO:0004024; P:alcohol dehydrogenase activity, zinc-dependent; IEA.  
DR GO:GO:0005499; P:electron transporter activity; IEA.  
DR GO:GO:0008270; P:zinc ion binding; IEA.  
DR GO:GO:0006118; P:electron transport; IEA.  
DR InterPro: IPR002328; ADH\_zinc.  
DR InterPro: IPR002085; Adh\_zn\_family.  
DR InterPro: IPR000345; Cytochrome\_B5.  
DR Pfam: PF00107; ADH\_zinc\_N\_1.  
DR PROSITE: PS00059; ADH\_ZINC\_1.  
DR PROSITE: PS00190; CYTOCHROME\_C\_1.  
SQ SEQUENCE 360 AA; 38933 MW; 06ED3B0581785759 CRC64;

Query Match 49.8%; Score 876; DB 10; Length 360;  
Best Local Similarity 48.9%; Pred. No. 1e-71;  
Matches 171; Conservative 52; Mismatches 77; Indels 50; Gaps

QY 1 ETGATGRFKVLYGVCHSDIHMAKDWGTSTPYTPGHGLVGVTVCCKVKPKSWRQ 60  
Db : |||||:::|||||:::|||||:::|||||:::|||||:::|||||:  
34 DNGENDWTVKLFCGVCHTDLTITKDWDGYSYFVPWGHEIVGLIATKVGNVTKFEKGSR 93  
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:  
QY 61 -----GCWLHGRLRPTCENCIHLENYCPLIIQTITYGSKYVDGMTYGGYSNNMVIDE 113  
Db :|||||:::|||||:::|||||:::|||||:::|||||:::|||||:  
94 VGVGVISSGSC-----QSCECDODLENYCPQMSEFTYNAGSOTKNYGVSINIVDQ 146  
QY 114 HFIVRIPNLPDLGAAPLLCAGITTSPRWYVLGXPGXHLGWVEWRFRSRPPILNPLGI 173  
Db ::|||:FPNLPSDSGAPLLCAGITTVSPMKYYGMTEAGKHLCV-----ACL 192  
QY 174 WGSRLQSIAPPFLIKEGSGYSTPALMH-----LLRTDDOMEAAAMST 216  
Db |::::::::::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
193 GG-----LGHWAVWKGFALGKVVTSSTKAEEAINHLAGDSLVTTFDPKMKAIGT 247  
QY 217 MDGIIDTPAVVRPLEPIISLKTNKKVTVGVIAVOPLDFPELFIIRKWVAWSAIGMMK 276  
Db :|||:DIISAVALYPPLLGLLVNKGLIALGLPEKPDELMPFELVLRKWWGSDVGSMK 307  
QY 277 EQEMIDFAEHNTADIEVIPDYLNTHMERRVKKOVRFRFVIDENTL 326  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:  
Db 308 ETOMLDPCAKHNITADIETFKMEINTAMERLAKSDEVRYRFVINVANSL 357

RESULT 8  
O65621 PRELIMINARY; PRZ: 363 AA.

ID AC O65621;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Cinmamy alcohol dehydrogenase-like protein, subunit A (Cinmamy-  
DE alcohol dehydrogenase)-like protein, LCADA' (Putative alcohol  
DE dehydroxanase) (EC 1.1.1.195).  
GN LCADA OR F20D10.90 OR AT4G37970 OR CAD6.  
OS Arabidopsis thaliana (Mouse-ear cress);  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eucotids; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

01-AUG-1998 (TrEMBLrel. 07, Created)  
DT  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT  
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DT  
DT Cinnamyl alcohol dehydrogenase-like protein, subunit A (Cinnamyl  
DE alcohol dehydrogenase-like protein, LCAD<sub>A</sub>) [Putative alcohol  
DE dehydrogenase] (SC I.I.1.135)  
GN  
LCAD<sub>A</sub> OR F20D10.90 OR AT4G37970 OR CAD6.  
OC  
OC Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; ros  
OC eurosidia II; Brassicales; Brassicaceae; Arabidops  
OC



```

242  TMDFIIDTVAEHAALLPFLSLKYSKGLVALGILEKPLDLPFLVLRKKVGGSGIQGM 301
276  KETQEMIDTDAEHNITDIPIDYDINTAMERVVKDVRFRVIDVENTL 326
302  KETQEMLEFCAKHKIVSDIELIKMSDINSAMDRLVKSDVRYRFRVIDVANS 352

RESULT 11
Q9SJ25
ID Q9SJ25 PRELIMINARY; PRT; 376 AA.
AC
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 25, Last annotation update)
DE Cinnamyl alcohol dehydrogenase-like protein (EC 1.1.1.195).
GN Ar2G21730 OR CAd2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
[i]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Eueli C.R., Ketchum K.A., Lee J.-J., Rensing C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vancken S.E., Umayam L., Talbot L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White C., Bisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana.";
RT Nature 402:761-768 (1999).
RL [2]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
SEQUENCE FROM N.A.
RP Kim S.J., Kim M.R., Bedgar D.L., Moinuddin S.G.A., Cardenas C.L.,
RP Davin L.B., Lewis N.G.;
RA "Functional Reclassification of the Purative Cinnamyl Alcohol
RA Dehydrogenase (CAD) Multigene Family in Arabidopsis.";
RT Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
CC -I- COFACTOR: ZINC (BY SIMILARITY).
DR EMBL; AC007019; AAC20393.1; -.
DR EMBL; AY302077; AAP59430.1; -.
DR FIC; E84604; E84604.
DR GO; GO:0004024; F.alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016491; F.oxidoreductase activity; IEA.
DR GO; GO:0008270; F.zinc ion binding; IEA.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF001107; ADH_zinc.N; I.
DR PROSITE; PS00059; ADH_ZINC; I.
DR Metal-binding; Oxidoreductase; Zinc.
SQ
SEQUENCE 376 AA; 40909 MW; E85BBD9DFD9FIEA94 CRC64;

```

```
QY 113 EHFIVRIPDNLFDGAAPLLCAGITTYSPWYYGLDK-PCMHGVLGVWPRFRFRSRPPLNLP 171
Db 141 HRFVLSIPDLGSDGAPLLCAGITTYSPKYYGMYTKSGKRLGVN----- 186
QY 172 GLWGRSLQSLVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAM 214
Db 187 GLGG-----LGHIAVKGKAFGLRVTVISRSSEKEREADRLGADGFLVTTDSQKMKEAV 241
QY 215 STMGGIIDVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGG 274
Db 242 GTMDFIIDVSAHALLPLFSLKVNKGLVALGLPEKPLDPLFSLVGLGRKMWGSGQIGG 301
QY 275 MKETQEMIDFAAEHNITADIEVIPIDYLNAMERVVKDVRFRFRFVIDVENTL 326
Db 302 MKETQEMLEFCAKHKIVSDIELIKMSDINSMDRLAKSDVRFRFVIDVANSL 353

RESULT 12
QYX93
ID Q7XE98 PRELIMINARY; PRT; 420 AA.
AC Q7XE98;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Putative cinnamyl alcohol dehydrogenase.
GN OSUNBA0073D04.1.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017097; AAP53892.1; -.
SQ SEQUENCE 420 AA; 43979 MW; 11204EF1143288A9 CRC64;

Query Match 44.9%; Score 790; DB 10; Length 420;
Best Local Similarity 46.9%; Pred. No. 9.6e-64;
Matches 161; Conservative 56; Mismatches 88; Indels 38; Gaps 5;

QY 2 TGATDVRFKLYCGVCHSDIIMAKNDMGTSYPIVPGHGLGVNTEVCCKVKPSW-RQ 60
Db 92 TGGDDVAIKILFCGICHSDLCIKNEWKHSIYPLVPGHEIAGVTVGKVTFRKAGDRV 151
QY 61 GRCLHGLRPTCNCIHLENYCNLTQYSGKYDGTMYGYSNNMTDEHFIVRIP 120
Db 152 GVGCMWNSCR-SCSSCNNGFNHCPEGVFTYNSVDKDTYVYGSYSSVWVHERFVWFP 210
QY 121 DNLPLDGAAPLLCAGITTYSPWYYGLDKPCMHGVLGVWPRFRFRPPLNLPGLWGRSLQ 180
Db 211 EAMPLDVGAPLLCAGITTYTPMKYHGLNAPKRGV-----LGLGLG----- 253
QY 181 LVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAMSTWDGIIDT 223
Db 254 --HVAVFAFAFGUKVTVISSPGKKEALERLADAPVWSSSABEMEAEASTWDGVNT 311
QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGGKETQEMID 283
Db 312 VSANTPMAPVIALKPKNGKMLVGLPENPLEVFPFSLVGHGRTLAGSNGGMAQTQEMIE 371
QY 284 FAEHNITADIEVIPIDYLNAMERVVKDVRFRFRFVIDVENTL 326
Db 302 MKETQEMLEFCAKHKIVSDIELIKMSDINSMDRLAKSDVRFRFVIDVANSL 353
```

```
Db 372 LAKHGVTADIEVIGADDVNTAMERLAKADVRFRFVIDVGNLT 414

RESULT 13
QYX722
ID Q9M722 PRELIMINARY; PRT; 289 AA.
AC Q9M722;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ELI3 (Fragment).
GN ELI3.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Rio Grande;
RA Thimmony R.L.; Martin G.B.;
RT "Rapid Induction of a Novel Gene and an eli3 Homolog During the Pto-
mediated Resistance Response in Tomato."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF146691; AAF72100.1; -.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002085; Adh_zn family.
DR Pfam; PF00107; Adh_zinc_N; 1.
FT NON TER 1
SQ SEQUENCE 289 AA; 31333 MW; B2BAFDC316A66CE8 CRC64;

Query Match 44.6%; Score 784.5; DB 10; Length 289;
Best Local Similarity 52.4%; Pred. No. 1.8e-63;
Matches 161; Conservative 50; Mismatches 57; Indels 39; Gaps 6;

QY 33 GHELVGVVTEGCKVKKFK-SWRQRCWMLHGLRPTCNCIHLENYCNLTQYSGKY 96
Db 1 GTRVVGWVTEGSKVKFKGDKGVGCMVSGCR-KCENCSDLENYCPRQIPTNYGSL 59
QY 97 DGTMYGYSNNMTDEHFIVRIPDLGGAAPLLCAGITTYSPWYYGLDKPGMHLGV 156
Db 60 DGTLTFGYSDVMVSDHEFVWVWPNLSMD-AAPLLCAGITTYSPKYPGLDKPGMHLGV 118
QY 157 EWRFRFRSRPPLNLPGLWGRSLQSLVPLIIGRKWVAGSAIGGSAIGGSAIGG 231
Db 119 -----VGLGG-----LGHMAVFAFAFGTKVTVISTANKKKEALERLGAJ 159
QY 202 --LLRTDQDQMEAAAMSTWDGIIDTVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVP 259
Db 160 SFLISRDPEQWKAAMNTLDGIIDTVSVVHPLILPLMLMKSHGKLWVWGAPEKEVELPVP 219
QY 260 LIIGRKWVAGSAIGGKETQEMIDFAAEHNITADIEVIPIDYLNAMERVVKDVRFRFR 319
Db 230 LLMGRLVAGSICGGMKETQEMIDFAAEHNITPDIEVVPMEVNTALERLLKSDVKYRFV 279
QY 320 IDVENTL 326
Db 280 LDIGNTL 286

RESULT 14
QYXWU0
ID Q7XWU0 PRELIMINARY; PRT; 360 AA.
AC Q7XWU0;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSUNBA0065B15.11 protein.
GN OSUNBA0065B15.11.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

```

OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Ma J., Yu Z., Chen L., Fan D.L., Meng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Chen S.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Zhang X.T., Ni L., Zhu P.H., Hong G.F.;
RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL731598; CAD39907.1;
DR SQ SEQUENCE 360 AA; 39086 MW; 0DEA56D0DBB425CA CRC64;

Query Match 43.0%; Score 755.5; DB 10; Length 360;
Best Local Similarity 46.5%; Pred. No. 1.1e-60;
Matches 160; Conservative 45; Mismatches 96; Indels 43; Gaps 8;

QY 1 FTGATDVRPKVLYCGVCHSDIHMKNOWTSTYPIVPGHGLGWVTEVGCKVKKFK-SWR 59
Db 32 ENGVDVTLKVKYCGMCHTDLHFINDWGITPTYPVPGHEITGVTKVGTNAGFKVGR 91
QY 60 QG-RCHLGRLEPTCENCCHLENYCPNLQTYGSKYDGTMTYGGYNNMVTDEHFIVR 118
Db 92 VGVGCIASCL--DCBCHRRSENYCDKVALTYNGIFWDGSIITYGGYSGMLVAHKRFVVR 149
QY 119 IPDNLPLDGAAPLLCAGITTYSPWRYGL--DKPGMHLGVMPFRFRPPLNLPGL 175
Db 150 IPTLELDAAAPLLCAGITTYSPMKHGMQLQADAGRRLGV-----VGLGG 195
QY 176 SRLQSLVPLPIKEGSGYSG-----TSPALMHS-----LLRTDQDQMEAAAMSTMD 218
Db 196 -----LGHVAVKFGKAFGLHVTIVISTSPAKEREARENKADNFVSTQKQQAATRSID 250
QY 219 GIIDTVPAVRPLEPLISLTKTKGVTVGIAVQPLDLPFPLIIGRMVAGSAIGMKET 278
Db 251 YIIDVAATHSGPILELLKXVGLVGAPEKPELPSFLIFGKRTVSGSWTGMKKT 310
QY 279 QSMIDFAAHNITADIEVIPIDYLTAMERVVKKDVFRFVIDV 322
Db 311 QEMDQICGEHNTDIEIVSTDRINDALARNDVRYFVNV 354

RESULT 15
QY Q8H809 PRELIMINARY; PRT; 391 AA.
AC Q8H809;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative mannitol dehydrogenase (NAD-dependent mannitol
DE dehydrogenase).
DE QJ1743A09.18.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,
RA Currie J., Collura K.;
RT "Rice Genomic Sequence."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC105364; AAN05338.1;
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002328; ADH_zinc.

DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR000345; CytC heme_BS.
DR Pfam; PF00107; ADH_zinc_N_1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR SQ SEQUENCE 391 AA; 40917 MW; EE3D9F7BB543C55B CRC64;

Query Match 40.5%; Score 712.5; DB 10; Length 391;
Best Local Similarity 40.6%; Pred. No. 1.1e-56;
Matches 154; Conservative 55; Mismatches 93; Indels 77; Gaps 7;

QY 3 GATDVRPKVLYCGVCHSDIHMKNOWTSTYPIVPGHGLGWVTEVGCKVKKFKSWRQGR 62
Db 31 GDEDVWVKILFCGICHSDLSTIKNWCNKKYPPVPGHEIVGVVAEVSVAAPAA---GD 87
QY 63 CMLHGLRLPTCE---NCIHLENYCPNLQTYGSKYDGTMTYGGYNNMVTDEHFIVR 119
Db 88 TVGVGVIASTCRACRCDGFENYCGALVPSFNAALPDGATVHGGSFELAVVNORYVVR 147
QY 120 PDN-----LPDGAAPLLCAGITTYSPWRYGLDKPGMHLGVMPFRFRPPLNLPGL 173
Db 148 PGGGGGASAPLDRLAPLLCAGTVVYCPRLGLDRPGVELGV-----AGL 193
QY 174 MGRSLQSLVPLPIKEGSGYSGTSPALMHS-----LLRTDQDQMEAAAMST 216
Db 194 GG-----LGHVAVKFGKAFGVKVTIVISTSPWKEAVERLGADAFLLSTNAEQMKAAGT 248
QY 217 MDGIIDTVPAVRPLEPLISLTKTKGVTVGIAVQPLDLPFPL----- 260
Db 249 MDGIIDTVSAVHDLTLLTLRTHGQLVPGSPGKPVQLALYELQSGSPFVGPRIDWAGG 308
QY 261 -----IIGRMVAGSAIGMKETQSMIDFAAHNITADIEVIPIDYLTAMER 307
Db 309 PIRARWAVLMVCDVADGKSVAGSMIGMMDYQEMVDFAVEHGVAAVEVIGMEDTGNME 368
QY 308 RVVKDVRFRFVIDVENTL 326
Db 369 RLQKGDVRYFRVIDVANTM 387

Search completed: July 6, 2004, 13:38:47
Job time : 33.8458 secs

```



Creation date: 07-14-2004  
Indexing Officer: MMOHAMMED1 - MENEN MOHAMMED  
Team: OIPEScanning  
Dossier: 09857518

Legal Date: 07-09-2004

No.	Dccode	Number of pages
1	SRNT	26

Total number of pages: 26

Remarks:

Order of re-scan issued on .....